TRANSIENT RELAXATION OF DNA METHYLATION AT THE ONSET OF MEIOSIS

by

Valeriya Gaysinskaya

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

August 2016

Abstract

Meiotic prophase I (MPI) is a unique phase of the cell cycle, specific to germ cells and defining of sexual reproduction. MPI is a period of extensive and specialized homologous chromosome interactions and genetic exchange. Proper progression of MPI requires elaborate epigenetic control, deficiencies in which often lead to infertility. Changes in DNA methylation during MPI can endanger genome integrity by activating transposable elements (TEs) that when mobilized induce DNA breaks and mutations. Therefore, MPI was thought to be under strict surveillance by DNA methylation, whose levels were assumed to be high and stable throughout MPI. Interestingly, expression of LINE retrotransposons, specifically LINE-1 (L1)-encoded protein ORF1p has been observed in MPI germ cells of wild-type male mice. Since tight epigenetic regulation is associated with transposon silencing, we hypothesized that L1 expression in MPI may indicate relaxation of epigenetic silencing in meiotic germ cells. Thus, we investigated the dynamics of CpG DNA methylation during MPI. We enriched and isolated individual MPI stages by Fluorescence Activated Cell Sorting (FACS) and profiled individual MPI germ cells using whole-genome bisulfite sequencing, and RNA-sequencing. Using this approach we uncovered transient and stage-specific changes in DNA methylation dynamics. In contrast to the prevailing view, we show that male germ cells undergo genome-wide transient relaxation of DNA methylation (TRDM) during early MPI. Specifically, we find that a transition from pre-meiotic spermatogonia to meiotic onset in preleptotene spermatocytes is accompanied by genome-wide hypomethylation. Gradual, but uneven remethylation of the genome creates hypomethylated domains throughout meiotic prophase, with pre-meiotic levels of DNA methylation achieved only by late MPI. Our data are most consistent with a DNA replication-coupled mechanism of DNA demethylation in pre-meiotic S-phase. Intriguingly, a TRDM-independent set of hypomethylated domains emerges in mid to late MPI and is enriched in transcriptionally upregulated spermatogenic genes. Using *Mael -/-* mice defective in piRNA pathway, we show that early MPI offers an opportunity for TE expression and reactivation. We

demonstrate that if germ cells enter MPI with insufficient levels of DNA methylation at L1 elements, then during TRDM, meiotic onset can be hijacked to reactivate potentially active L1s. Cumulatively, we demonstrate that early MPI is epigenetically relaxed, exhibits dynamic DNA methylation pattern in MPI and that transient genome-wide DNA hypomethylation at meiotic onset might have implications in gamete quality control.

Thesis Advisor, First Reader: Alex Bortvin, PhD

Second Reader: Joseph Gall, PhD

Acknowledgments

In his "The Greatest Show on Earth," Richard Dawkins writes, that "evolution never happened by taking one adult form and coaxing it into the shape of another", but instead, "every adult grows as an embryo". My experience as a graduate student has indeed been one of evolutionary growth as a biologist and a scientist. It is during my graduate research that I most closely faced theory, experiment and the promise of discovery. For the opportunity to do such research, and for his generous support and guidance throughout the years I want to thank my mentor, Dr. Alex Bortvin. Alex is an ardent advocate of knowledge, a passionate story-teller and someone who places great emphasis on data interpretation and criticism – essential aspects of quality research. My work would not be possible without Alex's scientific insight, wit and imagination.

I thank my thesis committee members: Drs. Joe Gall, Allan Spradling, Jeff Boeke, Nick Ingolia Kasper Hansen, and Chen-Ming Fan for their support, advice and critical eye. I am grateful to Dr. Gall, for sharing his knowledge and wisdom and whose unparalleled gift of observation and sense of discovery are contagious. I am grateful to Dr. Spradling, whose scientific intelligence I have absorbed throughout my graduate studies and who, as a director of Carnegie's Department of Embryology during this time has created a generous, prolific and inclusive environment for experimentation and communication. I thank Dr. Hansen, who upon the recommendation of Dr. Ingolia, became my collaborator and guided my efforts in the bioinformatics analysis of differentially methylated regions.

I thank the administrative and Information Technology staff of Johns Hopkins University (JHU) and Carnegie Institution for Science, for their availability, kindness and commitment to student welfare. For their personal support I want to particularly thank Joan Miller, of JHU and Bill Kupiec of Carnegie, who frequently helped me with diverse matters. I thank Mahmud Siddiqi for his insight and help with light and confocal microscopy. I thank Eugenia Dikovsky for her

indispensible work; none of my experiments would be possible without the best practices of Eugenia and her team at the mouse facility. I want to thank Allison Pinder for her excellent skills in matters related to Sanger and next generation sequencing and maintenance of high-quality core sequencing facility at Carnegie.

I have learned and benefited from many faculty, research staff, students and postdocs at JHU and Carnegie. I want to especially thank Dr. Svetlana Deryusheva, with whom, for the past several years, I have had invaluable daily coffee breaks infused with substantive and enlightening discussions about, among other things, experimental detail and logic. I am very grateful for her above and beyond help and generosity, for sharing her scientific expertise and intelligence, her insightful criticisms and thoughtful editing of my writings. I am also very grateful to Fred Tan, who provided generous step-by-step guidance with all matters related to bioinformatics analysis, for sharing his computational skills and tricks and for helping me navigate the zoo of computational tools.

Mammoth thanks to my fellow lab mates who have encouraged and helped me in many ways - I have been fortunate to be associated with this group of co-workers: Dr. Godfried van der Heijden for being an exciting colleague and a fantastic rotation mentor and for setting the stage for my graduate work; Dr. Safia Malki, who has supported me from day one, and who has significantly helped me with countless lab things and aided thoroughly with preparation of talks, posters and manuscripts; Drs. Julio Castaneda and Pavol Genzor, two Bortvin lab graduate student alumni, for their encouragement and advice; Marla Tharp, the most junior graduate student in the lab, for her positive reinforcement, encouragement and very helpful last minute corrections of my manuscript and thesis. Brendan Miller, for a terrific collaboration – I am very grateful for his important computational contribution to my work; I want to thank our lab technician, Rejeanne Juste, for all her valuable help and care on the daily basis.

Finally, I want to thank my family; My mom Marina and uncle Lenny, for their boundless love, support and encouragement; My sister Lina, for unfailing encouragement, unparalleled support, enthusiasm, and successful attempts to understand my research; My friend Polina for her solid friendship and generosity. Most particularly, I want to thank my husband Leonard, who has sustained me with unfiltered love and support, perceptive, uninhibited and honest views on life and world affairs and whose extraordinary art captures my imagination every time and finds its way into my world of science.

Table of Contents

Lists of Figures

List of Tables

List of Protocols

CHAPTER 1

Introduction

1.1 An overview

Germ cells give rise to gametes – sperm in males and eggs in females. Successful fertilization of an egg by a sperm results in the formation of a single cell called zygote, which gives rise to an entire organism. The story of an organism, then, is as much a story of existence, as it is a story of the survival of the fittest germ cell. Indeed, germ cells are unique cells with special properties. Meiosis, the playfield for my thesis research, is perhaps the most distinctive germ-cell specific property of all. Meiosis is a specialized cell cycle that ensures the production of haploid gametes from diploid progenitor germ cells.

While genetic complexity of a cell provides a blueprint for development, epigenetic complexity provides a blueprint for developmental regulation. The word epigenetic, the literal meaning of which is "above genetics", can thus generate complexity "above" and beyond that generated by the genetic code. The epigenetic complexity of a cell is set by a finite range of epigenetic marks, which are defined as covalent modifications of the DNA or post-translational modifications of the histone proteins that package DNA tightly into chromatin. DNA methylation is a key epigenetic DNA modification in mammals and involves the addition of a methyl group on the carbon-5 position of the cytosine ring of DNA, predominantly in the context of CpG dinucleotides in mammals [1]. The dynamics of cytosine methylation and demethylation are well known to influence the course of the developing germ cells, in terms of gene expression, chromatin compaction and other aspects of cell biology and development [2]. Consequently, improper DNA methylation patterns in germ cells are associated with genomic instability and infertility [3, 4]. In my work, I identify, describe and begin to explore the significance of changes in DNA methylation in meiosis, specifically meiotic prophase I (MPI) of mouse male germ cells.

Section 1.2 The rise of male germ cells

In mice, epiblast cells give rise to primordial germ cells (PGCs), the precursors of oocytes and spermatozoa (Spz). PGCs initially form a cluster of 30-50 cells in the extra-

embryonic mesoderm on embryonic day 7.25 (E7.25) [5] (**Fig 1-1A**). These PGCs start to proliferate and migrate into the developing gonadal region (the genital ridge). Upon approaching the genital ridge, migrating PGCs become connected by cytoplasmic bridges [6] and at E10.5, a batch of ~1000 PGCs reach the developing gonads. Within the genital ridges, PGCs continue to proliferate, reaching about 26,000 cells by E13.5 [5]. Once PGCs fully colonize the gonads by E13.5, PGCs undergo sex differentiation and male primary testes and female ovaries can be distinguished from each other morphologically [7, 8].

At E13.5, male germ cells undergo cell cycle arrest at G1/G0 and do not enter meiosis during the embryonic stages of development, as female germ cells do [5]. At birth, mitotically arrested mouse male germ cells resume proliferation by postnatal days 1-2 (P1-P2) [9]. During the subsequent month, the first wave of postnatal spermatogenesis, which starts with spermatogonia (Spg), occurs in a juvenile mouse, resulting in a first batch of mature sperm in 31-33 days [10] (**Fig 1-1B**). Subsequently, each new spermatogenic wave in the adult mouse testis, results in a continuous production of sperm. Spermatogenesis is divided into three main parts, (1) the mitotic proliferation of Spg, (2) meiosis of spermatocytes and (3) spermiogenesis, or differentiation and maturation of round spermatids into mature sperm (**Fig 1-1B**). This entire process occurs within the seminiferous tubule of testis, where Spg are found at the basement membrane, but spermatocytes and spermatids are found progressively further from Spg and towards the lumen, where mature sperm will be released (**Fig 1C**) [11]. During Spg proliferation, a self-renewing pool of spermatogonial stem cells, single cells located at the basement membrane of the seminiferous tubule, divide to renew themselves and to give rise to the more advanced Spg (**Fig 1-1B**). The first several rounds of mitotic divisions give rise to undifferentiated Type A Spg and include a renewed A single (A_S) and more advanced A paired (A_{pr}) and A aligned (A_{al}) (**Fig 1-1B**) [12]. Ultimately, each of these Aal cells undergo more mitotic cell divisions to produce differentiating Spg, including types A_1, A_2, A_3 and A_4, intermediate Spg (In) and type B Spg

(**Fig 1-1B**). Type B Spg divide to produce two preleptotene (PL) spermatocytes. Thus, about ten generations of Spg are necessary to form PL spermatocytes from one precursor stem cell. PL spermatocytes are cells that represent the beginning of meiotic prophase I (MPI), and are often considered as cells in meiotic onset. PL cells are primary spermatocytes that replicate and enter prophase of meiosis. The MPI is followed by two successive rounds of chromosome segregation resulting in the production of four haploid spermatids. The spermatids undergo a long process of spermiogenesis, during which a dramatic morphogenesis transforms them into mature Spz (**Fig 1-1B,C**). In principle, differentiation of one spermatogonial type A_S stem cell could lead to the production of up to 4096 Spz (**Fig 1-1B**).

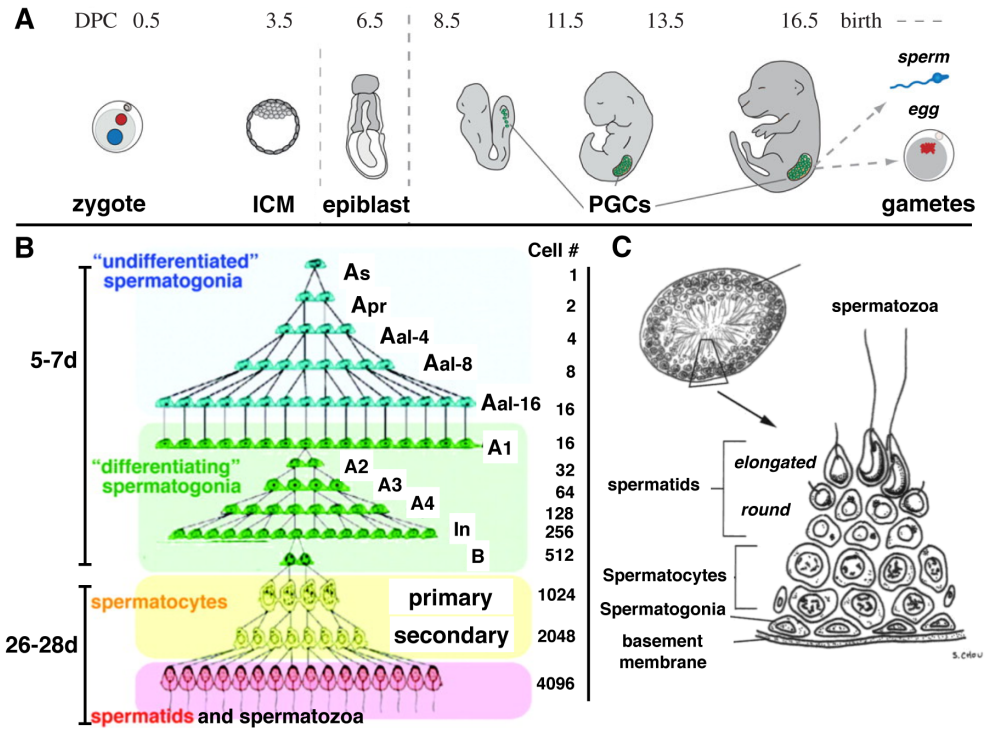


Fig 1-1. Diagram summary of gametogenesis and spermatogenesis. A) Gametogenesis overview. A zygote, which is a single cell that results from fertilization of an egg by a sperm, is made up of a maternal (red) and paternal (blue) pronuclei, but will give rise to an entire organism. After cleavage cell divisions, developmental milestones include the formation of the ICM, which will give rise to the embryo proper and to a more developmentally restricted epiblast. At around E7.5-8.5 (or 7.5-8.5 days post coitum, DPC), PGCs (green) arise from the epiblast, and eventually give rise to mature sperm and egg. **B)** Spermatogenesis overview, in males occurs starting 1-2 days after birth. In theory, one spermatogonial stem cell (As Spg) having undergone differentiation can give rise to 4096 spermatozoa. On the left of the diagram is the duration of spermatogensis in days, where it takes 5-7 days to undergo Spermatogonial mitotic divisions, and 26-28 days to undergo meiosis plus spermiogenesis. On the right of the diagram are values corresponding to the number of specific germ cell types. **C)** Diagram of testis cross-section. All spermatogonia are found at the basement membrane of seminiferous tubules. Spermatocytes,

round and elongated spermatids, are seen at successively higher levels within the seminiferous tubule. Mature elongating spermatids, also known as spermatozoa, are eventually released into the lumen.

Section 1.3 Germ cell development in the context of the mouse testis

Unambiguous identification of the exact stage of male germ cell development throughout spermatogenesis is a challenge, and an art. Classical criteria used for "staging" involves tissue fixation and PAS-hematoxylin staining, followed by identification of a particular stage (**Fig 1-2A**. e.g stage IV) of the seminiferous epithelium cycle, which is composed of the 12 stages (Clermont and Leblond, 1953; Oakberg, 1956; Russel et al 1990) (**Fig 1-2B**). Each of the twelve stages is characterized by the presence of specific cellular associations. For example, stage VI (**Fig 1-2A**) necessarily contains A and B Spg (or A and B-to-PL cells), plus P spermatocyte, plus "step 6" round spermatid and also, a more differentiated "step 15" elongated spermatid. Thus, any given stage will contain a unique combination of pre-meiotic Spg, meiotic spermatocyte(s) and post-meiotic spermatid(s) (**Fig 1-2B**). Recognition of stages depends on recognition of these cell types. It is, for example, extremely difficult to differentiate mid-P spermatocyte at stage VIII, from late-P at stage X. Much easier, is to recognize that a cross-section at stage VIII contains two types of spermatids (a round one at step 8 and an elongating one at step 16 of development), while a section at stage X, contains no round spermatids, but only an early elongating spermatid of step 10 (**Fig 1-2B**). Indeed, a sure way to differentiate stages is to differentiate the stages or steps of spermatid development. The development of an acrosome is a primary means to do so histologically (Russell et al, Hist and Histopath Evaluation of the Testis), where the development of the acrosome is accompanied by the dynamic changes in acrosome shape (**Fig 1-2C**).

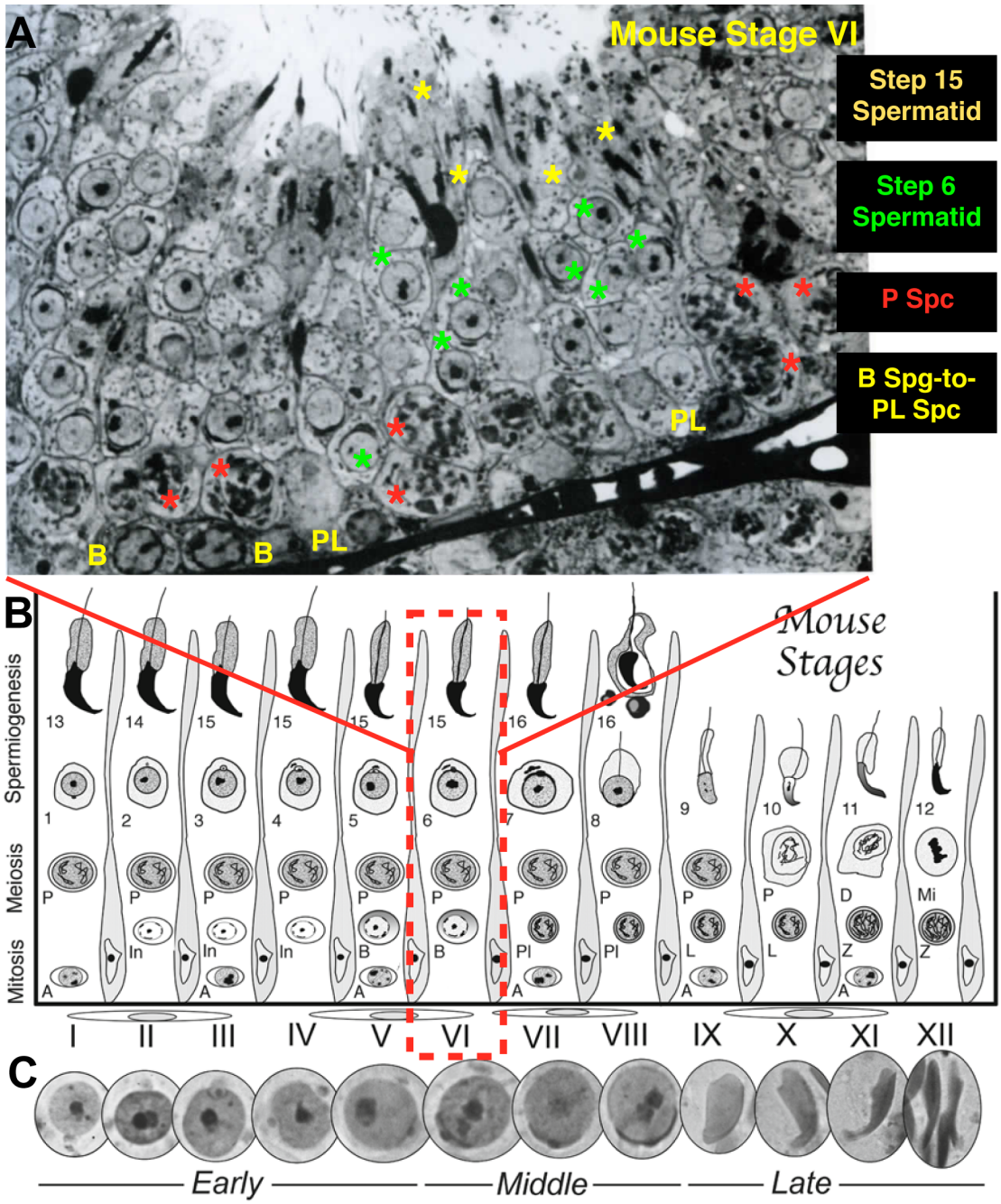


Fig 1-2. Spermatogenesis and the cycle of seminiferous epithelium. Spermatogenesis is a long process during which spermatogonial cells transform into spermatozoa. A) Identification of a particular stage of spermatogenesis, e.g. stage IV, is contingent on an ability to identify germ cell associations in any one testicular cross-section. Specific cell associations make up the cycle of the seminiferous epithelium. **B)** This cycle of spermatogenesis is subdivided into 12 morphologically

distinct stages in the mouse. Any given testicular cross-section is associated with one or more of the proliferative Spg cells, meiotic spermatogonia and post-meiotic spermatids. **C)** The early, middle and late development of the acrosome is the primary criterion for identifying stages of spermatogenesis. It is accompanied by the dynamic changes in the shape of the acrosome, from early proacrosomal granule, to flattening and spreading of the developing acrosome. In the earliest step 1 spermatid, no discrete acrosomal granules are typically observed.

Section 1.4 Meiosis

Meiosis, unique to sexually reproducing organisms, is a specialized cell cycle during which a single round of DNA replication is followed by two rounds of chromosome segregation and whose goal it is to reduce the chromosome number by half, from diploid to haploid. Thus, in sexually reproducing organisms, two parents will contribute one genome (one complement, or 1C) each to an offspring, so that the latter ends up with a 2C genome, containing two copies of the genome, the maternal and the paternal one. In mice, the very first meiosis commences at around 8-10 days after birth, with a new wave of meiosis starting every ~8.6 days throughout the animal's life [10, 13]. Prior to meiosis, type B spermatocytes, which are highly differentiated Spg committed to meiotic program, divide to produce PL spermatocytes, which undergo pre-meiotic DNA synthesis, resulting in the duplication of the parental genetic material, and beginning of meiotic prophase.

The soul of meiosis lies within the MPI during which, unique pairing of homologous chromosomes and recombination take place. MPI is exceptionally long, lasting about 3 weeks, during which time the chromosomes undergo dramatic changes in their organization (**Fig 1-3A**). MPI is subdivided into leptotene (L), zygotene (Z), pachytene (P), and diplotene (D) stages, based on their unique nuclear and chromatin morphology [11] (e.g., **Fig 1-3A**). Even though MPI is subdivided into distinct stages, a change from one stage to another involves rather gradual morphological transitions [11] that are poorly defined molecularly.

It is beyond the scope of this thesis to describe in detail all the events that make MPI unique, however, a reader is encouraged to visit the contents of in-depth literature on meiosis [14, 15] and other chapters of this dissertation. Below, several important meiosis-specific phenomena are reviewed with focus on meiotic entry, homologous chromosome pairing, recombination and stage transitions.

Meiotic S phase (or pre-meiotic S phase, used interchangeably) is much longer than mitotic S-phase in the same organism [16-18]. The difference in duration between mitotic and

meiotic S phases was found to be not due to the difference in replication fork movement rates, fork lengths or Okazaki fragment sizes [19]. Instead, it has been shown that considerably longer meiotic S phase is in part due to the reduced number of replication initiation sites [17, 20]. While the basis of the difference in meiotic S phase length is unclear, it is presumed to depend on the packaging of chromatin and is thought to be important for the unique events of meiotic prophase I [17, 18, 21]. Such events may be those required for or coordinated with the establishment of meiosis-specific chromosome features, like specialized sister-chromatid cohesion [18, 19]. It is well known that DNA topology and chromatin modifications exert influence on the replication timing, for example, inactive heterochromatin replicates late during S phase [22]. Thus, it is likely that meiotic and mitotic DNA topologies differ during S-phase and this difference may contribute to differential replication timing. Yet, another potential difference in replication timing, could be due to the difference in accuracy of DNA replication, under the hypothetical premise that more accurate replication requires more time [23]. Since meiotic cells have the responsibility of contributing their genomes to the next generation, it is conceivable, that DNA replication in meiosis is more accurate than that in mitosis. Intriguingly, more types of DNA polymerases are found in meiosis than mitosis, with different DNA polymerases sporting differential accuracy and speed of replication [24].

After chromosome duplication in PL, the homologous chromosomes start to pair along their lengths, based on homology. Meiotic pairing is a truly unique and dynamic process, in the context of which key meiotic events occur. One way to demonstrate the main aspects of homolog pairing and the temporally parallel meiotic events is to follow the assembly of a meiosis-specific, structure, the Synaptonemal Complex (SC) (**Fig 1-3A**). SC marks and mediates stable and progressive interactions of the homologs, eventually bringing the homologs in close juxtaposition [25].

The SC assembly starts early in MPI. In L stage, SC proteins, including SCP3 and SCP2 begin to assemble on chromosomes, indicating early stages of pairing. At this time, these proteins

microscopically appear as short, thin stretches called Axial Elements (AEs) along which, sister chromatids of individual homologs organize in loops attached at the AE core (**Figs 1-3A,B).** The initiation of homolog pairing is very poorly understood, but is thought to take place in L stage, with initiation of programmed meiotic double-strand breaks (DSBs) by SPO11, an evolutionally conserved type II topoisomerase-like protein [26]. However, it is likely that in mouse spermatocytes, and in some other organisms, homologous pairing initiation occurs prior to DSB formation, during pre-meiotic S phase in PL stage [27]. Indeed, the occurrence of SC, a marker of homologous pairing, is observed during pre-meiotic S phase in mice, and is not unique to mouse model [28, 29].

Homologous chromosomes become paired in Z, forming the SC at points of association. Thus, in Z, homologous chromosomes become closely associated by synapsis at various points along their lengths, with AEs, appearing as long thin stretches along the length of the chromosome (**Fig 1-3A**). In Z, AEs start to zip up via the laying down of the Central Element (CE) made up of proteins including SCP1 (**Fig 1-3B**). As chromosomes become paired along their lengths, they undergo meiotic recombination, which is an essential source of genetic diversity. Meiotic recombination (a physical exchange of chromosome segments) is initiated with the formation of programmed DSBs, but stable homologous recombination occurs in the context of the SC, starting at Z. Much of meiosis is dedicated to repairing DSBs, and meiotic homologous recombination is one form of repair. The L-to-Z (LZ) transition is truly unique, and is defined as the period when homologs, already extensively paired, come close together forming the bouquet, a meiotic chromosome configuration where chromosome ends cluster and "integrate" into the nuclear envelope in a polarized fashion, creating an organization reminiscent of a bouquet of flowers held at the stems (although in the case of the meiotic bouquet, both ends of the chromosomes cluster). LZ transition is marked by extensive chromosome movements, both passive and active. The Z-to-P transition (ZP) is marked by the completion of meiotic synapsis and recombination. In P, AEs, now called Lateral Elements (LEs), are fully zipped up by CEs,

marking the end of the SC formation, and indicating homologous chromosome synapsis. In P, but initiated at Z-to-P transition, meiotic sex chromosome inactivation (MSCI) and meiotic silencing of unsynapsed chromatin (MSUC) occur in male germ cells [30]. MSCI is a process of XY inactivation via transcriptional silencing and compartmentalization into a nuclear subdomain called the sex-or XY-body, that persists through diplonema. MSUC silences autosomes that fail to properly pair with their homologs. Pachynema, is an extremely busy stage of MPI, exhibiting a dramatic increase in overall RNA and protein synthesis in preparation for the next phase of spermatogenesis. At Diplonema, SC disassembles and homologous chromosomes de-synapse. By the end of the diplotene stage, the homologs remain connected only at chiasma, the physical remnants of homologous recombination.

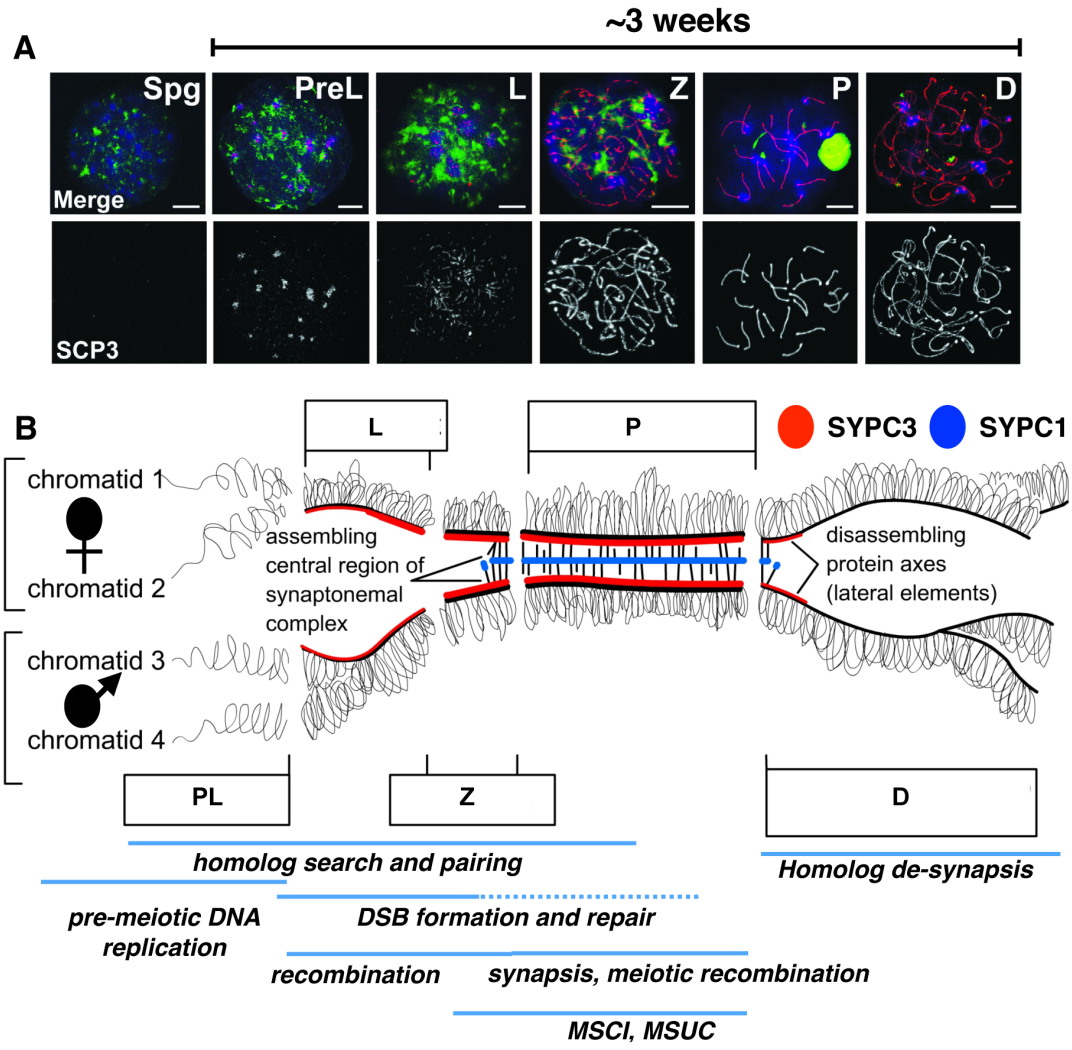


Fig 1-3. Key features of Meiotic Prophase I (MPI). A) Timing of MPI. Morphology of the nuclei of different stages of male meiosis by immunofluorescence analysis. Spg - differentiated spermatogonia; preL - preleptotene spermatocytes; L-leptotene spermatocytes; Z-zygotene spermatocytes; P - pachytene spermatocytes; D - diplotene spermatocytes. Spread nuclei were double labeled with γH2AX (green) and SYCP3 (red) and counterstained with DAPI (blue). Staging was deduced from changing, stage-specific labeling patterns of these markers. Bar - 10 μm. **B)** Diagram of SC assembly in MPI and summary of key MPI events.

Section 1.5 Epigenetic dynamics in gametogenesis and meiosis

Epigenetic changes are those that can influence the phenotype without changing the genotype, yet have an ability to be inherited. Key epigenetic modifications include DNA methylation and histone modifications. Changes in these modifications can lead to changes in transcriptional program of the cell, chromatin remodeling, perpetuation of a particular chromatin structure through generations and more. Germ cell development is accompanied by extensive epigenetic transitions. In the following paragraphs, several epigenetic changes that are critical for proper germ cell development are highlighted, with focus on meiosis, a hub of epigenetic plasticity and diversity.

Changes in Histone Modifications. Histones package and order genomic DNA into a structural complex called the nucleosome. Histone modifications affect nucleosome structure and stability and therefore chromatin organization. There are a large number of histone posttranslational modifications (PTMs), including histone methylation (primarily of lysines and arginines), histone acetylation (of lysines) and others. Histone PTMs play an important role in male germ cell development. While it is beyond the scope of this introduction to review histone PMTs, it is important to mention their significance in germ cell development. Major changes in histone PTMs occur in meiotic and post-meiotic germ cells. Below, several illustrative examples are provided.

Post-meiotic chromatin remodeling involves the removal of canonical nucleosomal histone proteins (upon histone H4 hyperacetylation) and their replacement with histone variants; first with the transition nuclear proteins and later, with basic proteins called protamines [31]. The histone-to-protamine exchange is essential for the extreme condensation of the genome in mature sperm.

Large-scale histone exchange is also important for meiotic spermatocytes. For example, MSCI is associated with chromosome-wide replacement of H3.1/H3.2 with H3.3 [32]. Similarly, shortly after MSUC induction in late Z/early P spermatocytes, H3.1/H3.2 nucleosomes are

evicted, and H3.3 incorporated into unsynapsed autosomal chromatin [32]. This loss of H3.1/H3.2 coincides with loss of most histone PTMs, with a few exceptions. Functionally, the progressive gain in H3.3 from late P-to-D spermatocytes is concomitant with increasing levels of autosomal transcription [33], because H3.3 attracts activating histone modifications conducive for transcription in euchromatin [34].

Dynamic mono-, di- or tri-methylated H3K4 (H3K4me1,2,3) and H3K9me2 modifications have been observed in MPI [32, 35]. The functional importance of these changes is in infancy of elucidation, but the progress is exciting. In MPI germ cells, H3K9me1,2 is widely localized to euchromatin of early L through late Z spermatocytes, disappearing in pachynema [35] and reappearing in heterochromatic foci of late P and D cells [32]. H3K9me1,2 modifications are known to contribute to transcriptional silencing. Mutant mice in which an H3K9me1,2 histone methyl transferase (HMTs), G9a, is mutated, exhibit an overexpression of inappropriate genes in meiosis, disrupting the gene silencing crucial for MPI progression [35]. In another study, transient acquisition of H3K9me2 at pericentric heterochromatin of spermatocytes was found crucial for facilitating homologous chromosome interactions via maintenance of their centromeres in close proximity [36]. In general, loss of H3K9 HMTs negatively affects pairing of the homologous chromosomes in mouse MPI, and is associated with delayed synapsis and impaired post-meiotic development [37].

Histone methylation also plays a very significant role in meiotic recombination. In mammals, crossing-over involves repair of meiotic DSBs, which map to preferred chromosomal locations called hotspots. Recently, it was demonstrated that meiosis-specific HMT that mediates H3K4me3, PRDM9 (also called MEISETZ), binds to the hotspot motifs [38]. Meiotic recombination is thus thought to be initiated when PRDM9 deposits H3K4me3 and triggers SPO11 mediated DSBs.

Changes in DNA methylation. An essential form of epigenetic change is the methylation of cytosine at its 5' position (5meC). 5meC in DNA of mammalian cells is found almost entirely within CpG dinucleotides [39]. A great deal of attention has been focused on the endogenous patterns of CpG DNA methylation. This is because this epigenetic mark has significant regulatory and structural roles in a variety of fundamental developmental processes, such as genomic imprinting, where DNA methylation is essential for regulation of parent-specific gene expression [40, 41]. One critical function of DNA methylation is transcriptional regulation, although the relationship between DNA methylation and gene transcription is complex. At CpG-rich gene promoters, DNA methylation promotes and directs gene silencing [42, 43]. Transcriptional repression, associated with low promoter DNA methylation, is in part, mediated by the methyl CpG binding proteins (MeCP) and methyl domain binding proteins (MBD) [44]. The reverse relationship, that of gene expression and undermethylation is not as straightforward, since CpG-rich regions (called CpG Islands, or CGIs) within the promoters of many genes are typically unmethylated regardless of the transcriptional status of the genes [44]. However, there are many exceptions, one, being the CpG-rich promoters of transposable elements (TEs) that are normally methylated, and whose DNA demethylation is associated with their transcriptional activation [4, 45]. The role of DNA methylation in TE regulation is discussed later in the chapter.

Mammalian development is characterized by bimodal DNA methylation reprogramming that occurs first during pre-implantation (with the lowest levels, of < 20% methylation genome-wide, in the inner cell mass at ~E3.5) and then during embryonic germ cell development (reaching lowest DNA methylation levels, of < 20% genome-wide, at E12.5-E13.5) (**Fig 1-4**) [46-48]. The former DNA methylation reprogramming begins in the zygote, when, upon fertilization, rapid demethylation of the paternal genome (male pronucleus) is observed [49]. Additionally, in the first few zygotic cleavage divisions, DNA methylation of both maternal and paternal genomes gets diluted in a replication-dependent manner, reaching its lowest levels in the ICM of the pre-implantation embryo. DNA methylation is immediately regained post-implantation [50, 51] (**Fig**

1-4). In the event of PGC DNA methylation reprogramming, prior to reaching the genital ridges, PGCs exhibit parent-of-origin-specific imprinting methylation marks (called genomic imprinting), which enforce the mono-allelic expression of many imprinted genes. Most parental methylation imprints on paternal and maternal alleles are erased between E11.5 and E13.5 [52] and this genome-wide erasure of parental DNA methylation marks is associated with acquisition of totipotency. The global DNA demethylation is followed by the de novo methylation that involves germ-cell specific resetting of DNA methylation marks, which occurs only after sex determination, depending on sex of the developing embryo (**Fig 1-4).** De novo methylation takes several days, but is largely completed by birth in the male [53].

DNA Methyltransferase enzymes (DNMTs) are responsible for creating and propagating all DNA methylation patterns. DNMTs catalyze either *de novo* methylation (DNMT3a, DNMT3b) or maintenance methylation (DNMT1) of hemi-methylated DNA following DNA replication [54, 55]. Dnmt3L, while lacking DNMT enzymatic activity, is critical for stimulating the activity of other DNMTs, like DNMT3a [55, 56]. These enzymes play a critical role in the genome-wide DNA methylation events that take place throughout germ cell development and their loss is detrimental to germ cell survival [3]. Different DNMTs are dynamically expressed throughout male germ cell development (**Fig 1-5**). The Dmnt3a and Dnmt3l peak during embryonic DNA methylation reprogramming, but their expression is heavily downregulated prior to meiosis [56]. While relatively lowly expressed during embryogenesis, Dnmt3b and Dnmt1 mRNA exhibit a modest, transient increase after birth, but are down-regulated in meiosis. The highest DNMT1 protein levels in postnatal testes were found in proliferating mitotic Spg, with gradual decline in meiosis.

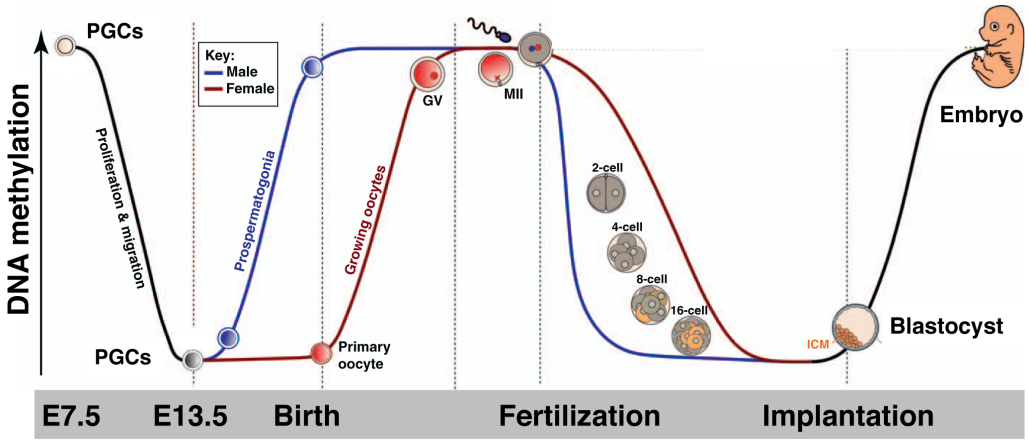


Fig 1-4. DNA methylation changes during embryogenesis and gametogenesis. Proliferating and migrating germ cells undergo epigenetic reprogramming, whereby DNA methylation is globally erased. E12.5-E13.5 marks the lowest point of methylation in an embryo. After sexual differentiation, mitotically arrested prospermatogonia regain methylation, and by birth, progenitor germ cells are highly methylated. After sexual differentiation, females take a different route and enter meiosis, whereby their levels of DNA methylation remain low. Post-fertilization, another DNA demethylation event takes place and is distinct on parental genomes. DNA demethylation at male pronucleus is fast and slower in female. Preimplantation embryo has very low levels of DNA methylation. Concomitant with implantation, the blastocyst undergoes remethylation. New methylation patterns are established according to the cell type and cell differentiation program.

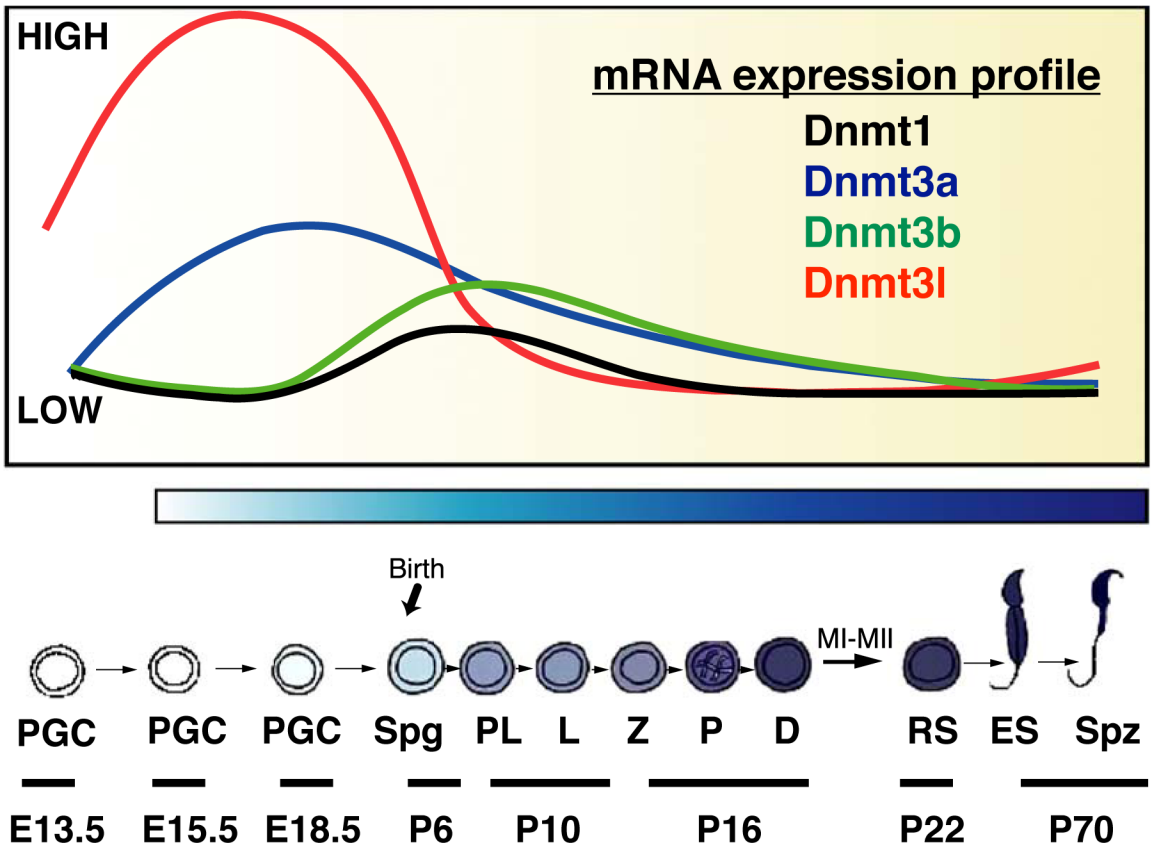


Figure 1-5. Diagram of DNA methyltransferase (Dnmt) dynamics in gametogenesis. Relative mRNA levels of Dnmt1, Dnmt3a, Dnmt3b and Dnmt3l are shown. The data is based on the real-time RT-qPCR from testis corresponding to embryonic stages (E13.5, E15.5 and E18.5) or postnatal stages (P6, P10, P16, P22 and P70). These ages correspond to embryonic DNAme reprogramming and after birth, starting with P6, to first wave of spermatogenesis. Approximately, the postnatal testis are enriched in mitotic Spg at P6, PL and L at P10, Z, P and D at P16, round spermatids (RS) at P22 and and elongating spermatids (ES) and mature spermatozoa (Spz) at P70.

Section 1.6 Transposable element silencing in the germline: DNA methylation and piRNA pathway

First discovered in maize by Barbara McClintock in the 1940s, transposable elements (TEs) are genomic DNA sequences that have the ability to become mobile and move from place to place in the genome. Due to their mobile nature, McClintock, who described TEs as "controlling elements", showed they can cause chromosomal breaks and mutations [57]. It is now a common knowledge that TEs make up the majority of the mammalian genome [58, 59]. We also know, that only a small subset of TEs is potentially mobile or active, with the majority of TEs mutated and thus benign in terms of their capacity to autonomously mobilize. Hence, the TEs in the genome can be divided into inactive and potentially active elements and correspondingly, into old and evolutionarily young ones [60].

There are many different types of TEs, and they are extremely diverse in sequence and structure. TEs can be categorized into two well-defined classes according to their structure and movement strategies. Class II elements, which only make up about 3% of the mouse genome, are made up of DNA transposons that are mobilized by a 'cut and paste' mechanism [61]. Class I elements, which account for almost half of the genome in mammals, mobilize by a 'copy and paste' mechanism involving an RNA intermediate [62]. A majority of these retrotransposons are truncated and mutated, but some active elements are capable of mobilization.

TEs are highly interspersed throughout the genome, and have played a major role in shaping the mammalian genome structure, gene expression and epigenetic topology of chromatin. Indeed, besides contributing large amounts of DNA to the mammalian genomes (including at least 40% of human and mouse genomes), TEs have significantly contributed to the repertoire of gene expression, providing new genes [63], exons [64, 65], and regulatory motifs [66]. Indeed, new TE insertions continue to impact and diversify gene expression [67].

Several families of retrotransposons co-exist in the genome. Retrotransposons are classified into long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and long

terminal repeats (LTRs). Long INterspersed Element-1 (LINE-1 or L1) is the major category of TEs in placental mammals, including mice, and contributes to over 20% of genome size, with ~500,000-600,000 copies of L1 in humand and mouse genomes [62, 68]. Most retrotransposons are mutated "fossils" of previously active TEs, but L1 superfamily hosts a fraction of active elements, with at least 3,000 potentially active elements [69]. Given the activity of LINEs in mouse genome, L1-dependent SINE activity is also observed in the mouse, with, for example, more than 300,000 SINE element insertions estimated since mouse and rat diverged [70]. LTRs also contain active copies in the mouse (and human) genome, with most mutagenic insertions from the intracisternal A particle (IAP) and MusD/ETn elements. Currently, there are multiple potentially active L1 families, distinguished primarily, but not exclusively, by the type of promoter they have [60]. LINEs are strongly biased towards AT-rich regions, whereas a major inactive category of retrotransposons, short interspersed nuclear element (SINE) is biased towards GC-rich regions, preferences that are also evident in the human genome.

Although a rich source of evolutionary variation, L1 activity can be detrimental to the fitness of the host, either by inserting into the genome, deleting and rearranging DNA or otherwise modifying the genome. Indeed, the role of TEs in human diseases is widely appreciated, from early discovery that *hemophilia* A resulted from a de novo insertion of L1 [71] to a growing list of diseases caused by TE-mediated mutagenesis. Most recently, TE expression has been associated with sporadic cancers or exacerbation of cancer phenotypes [72, 73]. Much of the TE activity, in both humans and mice, is due to active L1 retrotransposons.

To ensure their propagation, TEs have to mobilize within germ cells, the cells destined for the next generation. Indeed, during epigenetic reprogramming of the embryonic germline, when relaxation of epigenetic control is observed, TE expression is also observed and constitutes a dangerous window in which TEs can potentially escape from host restraint, mobilize and expand [74]. However, embryonic germ cells are equipped to combat TE expression. As mentioned earlier, DNA methylation is critical for transposon control, and DNMTs play essential

roles in the establishment and maintenance of retroviral methylation in embryogenesis (**Fig 5**). Since the damaging effects of TEs could jeopardize germ cell genome integrity and germ line propagation, in addition to DNA methylation, germ cells possess other defense systems. An essential line of defense includes the post-transcriptional RNAi-like control provided by the piRNA pathway [3, 75] (**Fig 1-6**). DNA methylation and the piRNA pathway are intimately linked at the level of TE control. Essentially, in collaboration between DNA methylation and the piRNA pathway, in what is referred to as the piRNA-directed DNA methylation pathway, TE silencing is ensured until the end of spermatogenesis [76, 77].

In addition to DNA methylation and piRNA pathway, various repressive histone modifications lock L1 and other TEs in the heterochromatic state [3, 75]. Indeed, there are a number of key repressors of TE activity in the male germ line whose inactivity leads to de-repression of TEs and germ line demise [3]. Key among these known repressors are components of DNA methylation machinery DNMT1, DNMT3a, DNMT3L, H3K9 methylatranserases ESET (also known as Setdb1 or KMT1E) and Suv39h1 (also known as KMT1A), and piRNA pathway components MILI, MIWI2 and MAEL [3].

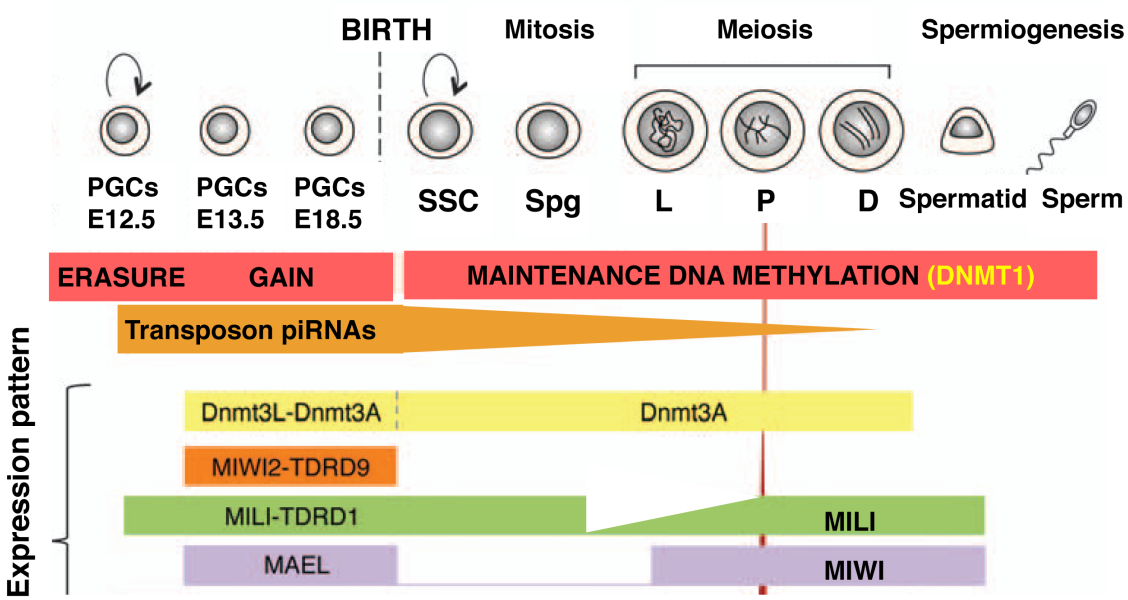


Fig 1-6. TE silencing in male germline. Genome-wide loss of DNA methylation during embryonic reprogramming is immediately followed by gain in DNA remethylation executed by DNMTs DNMT3a and DNMT3L. TE control also involves degradation of transposon RNA to transposon piRNAs by piRNA pathway, which includes MIWI2, TDRD9, MILI, TDRD2 and MAEL proteins. DNA methylation during meiosis is thought to be stable and maintained by DNMT1, with DNMT3A also available. Meiotic germ cells express some of the piRNA components, with early meiotic germ cells having seemingly low piRNA protein abundance.

Chapter 2

Flourescence Activated Cell Sorting (FACS) of murine spermatocytes

Chapter 2.1 Optimizing cytometry isolation of murine spermatocytes

Valeriya Gaysinskaya[1,2], Ina Y. Soh[1,2], Godfried W. van der Heijden[1], and Alex Bortvin[1,*]

[1]Department of Embryology, Carnegie Institution for Science, Baltimore MD 21218.

[2]Department of Biology, Johns Hopkins University, Baltimore MD 21218

*Correspondence: Department of Embryology, Carnegie Institution for Science, 3520 San, Martin Drive, Baltimore MD, 21218,

Phone: (410) 246-3044,

Fax: (410) 243-6311

E-mail: bortvin@ciwemb.edu

VG and IYS contributed equally to this work

Adapted and partially reprinted from *Cytometry A*. 2014 June ; 85(6): 556–565, entitled "Optimized flow cytometry isolation of murine spermatocytes".

doi:10.1002/cyto.a.22463.

Abstract

Meiotic prophase I (MPI), is an initial stage of meiosis characterized by intricate homologous chromosome interactions, synapsis and DNA recombination. These processes depend on the complex, but poorly understood early MPI events of homologous chromosome search, alignment and pairing. Detailed molecular investigation of these early events requires isolation of individual MPI substages. Enrichment for Pachytene (P) and Diplotene (D) substages of late MPI was previously accomplished using flow cytometry. However, separation of early MPI spermatocytes, specifically, of Leptotene (L) and Zygotene (Z) substages, has been a challenge due to these cells' similar characteristics. In this report, we describe an optimized Hoechst-33342 (Hoechst)-based flow cytometry approach for isolating individual MPI populations from adult murine testis. We get significant enrichment for individual L and Z spermatocytes, previously inseparable from each other, and optimize the isolation of other MPI substages. Our flow cytometry approach is a combination of three optimized strategies. The first is optimization of testis dissociation protocol that yields more consistent and reproducible testicular single cell suspension. The second involves optimization of flow cytometric gating protocol where a critical addition to the standard protocol for cell discrimination based on Hoechst fluorescence, involves a back-gating technique based on light scattering parameters. This step specifies selection of individual MPI substages. The third, is an addition of DNA content restriction to the gating protocol to minimize contamination from non-meiotic cells. Finally, we confirm significant enrichment of high-purity Preleptotene (PreL), L, Z, P and D MPI spermatocytes using stage-specific marker distribution. The technique will facilitate understanding of the molecular events underlying meiotic prophase I.

Introduction

Spermatogenesis, a multistage process by which spermatogonial stem cells differentiate into spermatozoa, occurs in the seminiferous tubules of the adult testis. As a result of asynchronous initiation of spermatogonial stem cell differentiation, the testis contains germ cells at all phases of

spermatogenesis. These include the proliferative phase, responsible for the mitotic expansion of diploid spermatogonia, the meiotic phase, which generates haploid cells from diploid progenitors by way of specialized two cell divisions preceded by one round of DNA replication, and spermiogenic phase, marked by the differentiation and maturation of haploid cells.

The cellular heterogeneity of the testis necessitates developing methods for germ cell enrichment and isolation. The study of individual stages of spermatogenesis has become more accessible with Hoechst-based flow cytometry approaches. Indeed, flow cytometry is commonly used to enrich for murine spermatogonial stem cells [78-80] and more recently, for late MPI spermatocytes [81-83]. On the other hand, enrichment for high-purity early MPI populations has remained difficult, thus, these cell types have largely escaped comprehensive molecular analysis. One challenge is that testicular cell heterogeneity is not proportional with respect to the percent of each cell type present. Due to the logarithmic nature of spermatogenesis, early cell types are proportionately underrepresented. A spermatogonial stem cell can give rise to up to 1024 primary spermatocytes and four times as many haploid spermatids, and the maturation of spermatid into a mature spermatozoon is one of the lengthiest phases of spermatogenesis. As a result, post-meiotic cells occupy the majority of the total testicular volume [11]. This makes it difficult to enrich for and study the less abundant and more transient early MPI populations. In fact, the purification of individual L and Z substages has not yet been accomplished to date. A major goal of this study was to achieve separation of individual L and Z spermatocytes.

Early MPI is critical for proper meiotic progression. Originally characterized in terms of changes in the chromosomal morphology [84], L and Z substages differ in a number of cellular processes. Leptonema, which follows Preleptonema where pre-meiotic S-phase occurs, involves extensive chromatin reorganization, telomere-led chromosome movement and telomere attachment to nuclear envelope. These are thought to facilitate homologous chromosome (homologue) interactions, alignment and pairing [84, 85]. At this time, meiotic “programmed” double stranded breaks (DSBs) form and Axial Elements (AEs) start to assemble along the paired

sister chromatids. During L to Z transition, a widely conserved telomere configuration known as meiotic bouquet can be observed, and is marked by telomere clustering at the nuclear periphery [86, 87]. In Zygonema, homologues begin to synapse as AEs assemble into synaptonemal complex (SC), a proteinateous structure that will eventually juxtapose the homologs along their entire lengths in Pachynema of MPI. The events observed in early MPI, including wide-range chromosomal movements, visible changes in the distribution of heterochromatin and assembly of SC, imply extensive changes in chromatin organization. Some molecular details of these early meiotic processes have been described in *C. elegans* [88, 89], *S. cerevisiae* [90], *S.* [91, 92] and other non- mammalian organisms. However, detailed molecular understanding of early meiotic events in mammals is still lacking. Clearly, detailed molecular and mechanistic studies, and the subsequent generation of appropriate mutants, are necessary to understand the scope of regulation and coordination of these events, necessary for proper meiosis. Thus, the isolation of homogenous populations of early meiocytes is essential for definitive molecular and biochemical studies of MPI.

Recently, refined Hoechst dye staining protocols and flow cytometric analyses, have allowed for an enrichment of several mouse MPI populations [79, 81]. These efforts have led to the much needed molecular and/or genetic analyses including examination of pre-meiotic, pre-DSB homolog pairing in PreL spermatocytes [27], nucleosome profiling at recombination hotspots in spermatogonia, PreL or mixtures of L/Z and P/D cells [83] and other studies [93]. However, questions that require the separation of L from Z remain unanswered, and even isolation of high-purity individual P and D populations remains a challenge, with publications resorting to analyzing mixtures of L/Z and P/D spermatocytes [83, 94]. This and a number of concerns prompted us to re-examine and optimize all major steps of the published Hoechst-based flow cytometry protocols, from cell dissociation to flow cytometric analysis. One concern was the preparation of a reliable testicular single cell suspension. There are a number of published protocols available for testis dissociation and Hoechst dye staining of adult murine testicular cells

[78-80, 82, 94-96]. While all these protocols provide an excellent foundation for preparation of testicular single cell suspension, a challenge, still, is to obtain suspensions of consistently high quality between the experiments. In addition, since the available protocols vary from one another in a number of aspects, including the medium and the duration of testicular tissue dissociation, the duration and concentration of Hoechst staining and other parameters, the resulting Hoechst-labeled testicular suspension profiles differ between the laboratories. Another concern is cell purity. Published literature largely fails to document the purity of sorted cells, and/or note the criteria for purity determination. Consequently, practical information such as the number of cells examined, the markers used for evaluation, and the type of contaminants observed is largely unavailable. There is also an ambiguity associated with cell sorting and collection parameters, with often incomplete reporting on important parameters, including the rate of sorting and the number of populations sorted at a given time.

Previous studies often used juvenile testes as the means of reducing cellular heterogeneity of the starting material and/or maximizing the yield of early MPI spermatocytes. Our method is optimized for the adult murine testis, which provides an enriched source of all MPI substages. An additional important reason for using adult versus juvenile testis in our analysis, is that the first round of spermatogenesis is thought to substantially differ from all subsequent rounds, and has been linked to differences in undifferentiated spermatogonia (Spg) [97] as well as associated with increased apoptosis [98, 99]. As the differences in the first wave of spermatogenesis may suggest differences in the first meiotic wave, this raises a question of whether first meiocytes are distinct from the subsequent ones. Furthermore, our approach should allow sorting from adult testes of mutant mouse lines deficient in various aspects of male germ cell differentiation.

In this report we present an optimized flow cytometric analysis of Hoechst testicular single-cell suspension that greatly improves the purity and reproducibility of the sorted spermatogenic cell populations from adult murine testis.

Materials and Methods

Materials, reagents and solutions for testicular cell suspension preparation

Materials: 15 ml conical tubes (BD Falcon, #352097); Shaking Water Bath (VWT, #89032); Disposable transfer pipet (VWR); 100 μm nylon cell strainer (BD Falcon, #352360); 40 μm nylon cell strainer (BD Falcon, #352340); 12×75mm tube with 35-μm cell strainer cap (BD Falcon, #352235). Reagents: Collagenase Type I (Worthington Biochemical, #LS004196) DNAse I (Sigma, #DN25-10MG); Gey's Balanced Salt Solution (GBSS) (Sigma-Aldrich); 2.5% Trypsin (10X) (Gibco, #15090-046); Hoechst 33342 (10 mg/ml solution in water) (Life Technologies, #H3570); Newborn Calf Serum (NCS) (Life Technologies); Propidium Iodide (PI) (1 mg/ml) (Sigma-Aldrich, #P4864). Stock solutions: DNAse I (1mg/ml solution in 50% glycerol was made from 552 Kunitz units/mg powder and stored at −20°C); Freshly prepared solutions: "Collagenase I/Dnase I" (Collagenase type I [200 U/ml] and DNAse I [5 μg/ml] in GBSS); "Collagenase I/Dnase I/Trypsin" (Collagenase type I [200 U/ml], DNAse I [5 μg/ml], and Trypsin (0.025%) in GBSS).

Testicular single-cell suspensions and Hoechst-33342 (Hoechst) staining

For protocol details, see Section 2.2 of this chapter. Testis dissociation was based on a recently described method [82] with modifications. Our final protocol was conducted based on six consecutive steps described below.

1. Testis digestion: After the removal of tunica albuginea, each testis was placed in a 15 ml conical tube on ice with 6 ml "Collagenase I/Dnase I" solution. The tube was sealed with parafilm, shaken in horizontal position at 150 rpm for 10 min at 35°C. The temperature and agitation speed were the same for all subsequent incubation steps. Halfway into the 10 min incubation, the testis were gently pipetted, up and down twice to help tubule dispersion. This and all other pipetting steps were done using disposable transfer pipets. By the end of this step, tubules appeared thin and dispersed.

2. Somatic cell removal: Tubules were allowed to settle for 2 min at room temperature (RT) by standing the tube vertically. The supernatant, enriched in interstitial testicular cells was removed, leaving just enough liquid to cover the settled tubules.

3. Seminiferous tubule digestion: 6 ml of pre-heated "Collagenase I/Dnase I/Trypsin" solution was added to the tube and the tubules were gently pipetted up and down 10 times. Halfway into the 25 min digestion period, 60 µl of 2.5% Trypsin was added, and the tubules were pipetted again 10 times. At the end of the incubation time, pipetting was repeated 10 times. The tubules appeared fragmented and solution dense with cells. The resulting suspension was passed through a 100 µm nylon cell strainer. At this point, a 100 µl aliquot was removed to be processed for cell counting and viability estimation (see Materials and Methods below) while the rest of filtered cell suspension was pre-stained with Hoechst dye (during step 4 below).

4. Pre-staining with 100 µg Hoechst dye: To the resulting filtered cell suspension, 10 µl of DNAse I (1 mg/ml) and 10 µl of Hoechst dye (10 mg/ml) were added. The suspension was pipetted up and down 10 times and incubated for 20 min. Halfway into the 20 min period, the suspension was pipetted again. At the end of incubation, 600 µl of NCS was added to inactivate trypsin and the suspension was pipetted up and down 5 times.

5. Staining with Hoechst: After determining the cell number, the suspension was spiked with 10 µl of DNAse I (1 mg/ml), and stained with Hoechst dye for the final 6 µg Hoechst/million cells. The suspension was pipetted up and down 10 times and incubated for 25 min. Halfway into the incubation period, the solution was pipetted again, and then once more at the end of the incubation. Finally, the suspension was passed through a 40 µm nylon cell strainer. The suspension was kept on ice until sorting, which usually proceeded within 30 min to one hour after the completion of step 5.

6. Staining with PI: Immediately prior to sorting, 2 ml cell suspension was removed into a 5 ml polypropylene culture tube, the cells were stained with 10 µl of PI at room temperature and filtered into a tube with 35 µm cell strainer cap.

Cell Counting and Viability Estimation

Prior to Hoechst staining, viability and total yield of testicular single cells were determined. For trypan blue exclusion test of cell viability, 100 µl of the cell suspension was combined with 200 µl GBSS and 300 µl 0.4% trypan blue and analyzed using hemocytometer. Cell viability was found to be more than 95% in all cases. The total number of alive cells per testis (50 to 90 million cells depending on age) was calculated based on the total viable cells per ml. Note, that these counts underestimate the total number of cells per testis, since the counts were performed after filtration through 100 µl nylon cell strainer, a procedure that eliminates an uncertain number of elongated spermatozoa. Additionally, small size and the mobility of sperm cells introduce some difficulty to accurate sperm quantification, potentially further underestimating cell counts.

Prior to cell sorting, cell viability of the ready-to-sort samples was examined based on staining of cells with PI described above. The total cell viability, based on the number of PI-positive cells, ranged between 86 and 94 percent (Supplementary Fig 2-1).

Flow Cytometric Analysis and Fluorescence Activated Cell Sorting

For details, see Section 2.2 of this chapter. Data analysis was done using BD FACSDiva software. Hoechst was excited using 375 nm laser, and the dye's wide emission spectrum detected in two distinct channels: the "Ho Blue" (450/40 nm band-pass filter) and the "Ho Red" (670 nm long pass filter). The latter was also used to detect PI. A dichroic mirror (610 nm long pass filter) was used to split these emission wavelengths. Forward Scatter (FSC-A) and Side Scatter (SSC-A) were detected using 488 nm laser. Two-way sorting was performed using a seventy-micron nozzle size. Sorting Flow rate was adjusted to 2000–3500 events/second. A minimum of 500,000 events, were pre-recorded before setting of the gates. Cells were sorted into 5 ml polypropylene round-bottom tubes coated with (by pipetting) and containing 1 ml of 5% NCS in GBSS.

Meiotic spreads and Immunostaining

To estimate percent purity of individual sorted populations, an aliquot of each was processed for immunofluorescence staining. Nuclear spreads were prepared as described [100], with minor modifications. See Chapter 4 of this thesis (Methods and Procedures) for details.

Results and Discussion

Preparation of adult murine testicular cell suspension, Hoechst dye staining and flow cytometric analysis

A crucial first step towards successful flow sorting is preparation of a reliable testicular single cell suspension. Guided by the available published protocols, and predominantly by the one developed by Getun and colleagues [82], here, we optimize and present an improved testis dissociation and Hoechst dye staining protocol of adult murine testicular cells (detailed in the Materials and Methods section). The protocol has been optimized to produce a well-defined and consistent profile of MPI populations. We have incorporated a number of points in the testis dissociation and Hoechst dye staining procedures that facilitate the reproducibility of subsequent sorting results. For example, to minimize variability in cell dissociation efficiency, we performed all steps in a larger reaction volume and we have optimized the concentration of Collagenase I, DNAse I and Trypsin used. To avoid cell and DNA clumping, we periodically mix cells throughout the cell dissociation step and spike the testicular suspension with Trypsin and DNAse I. Perhaps the most critical optimized aspect of our protocol involves adjusting Hoechst concentration per total cell number counted per testis. We use 6 μg Hoechst dye per million cells, and depending on total cells counted, a different amount of Hoechst dye will be added to a particular testicular cell suspension. This is in contrast to, for example, adding a fixed amount of Hoechst dye per testis. Indeed, variability in Hoechst staining protocol easily leads to discrepancies in testicular Hoechst profile between the experiments. Although this protocol, like any other that involves Hoechst dye staining, exhibits sensitivity to variations that may influences

Hoechst equilibration and subsequent Hoechst-stained cellular profile [101], our protocol represents an improvement towards consistency and reproducibility between experiments. This point is illustrated by similar Hoechst profiles of different testicular digests (Supplementary Fig 2-2.)

Subsequently, Hoechst- and PI- dye stained testicular cell suspension is processed for flow cytometric analysis. It is essential to minimize noise from the unwanted sources such as debris. This is done on the basis of the cells' light scattering parameters, namely, the Forward Scatter (FSC) and the Side Scatter (SSC), proportional to the cell size and cell granularity, respectively. Since most debris particles are typically of small size, the debris is excluded by setting a gate that excludes signals with low FSC intensity (**Fig 2-1A**). Due to the very small size of elongated spermatozoa, exclusion based on low FSC also eliminates much of their contribution from the analysis and sorting.

Subsequent visualization of the selected cells simultaneously in Hoechst blue and red channels reveals a complex fluorescence profile (**Fig 2-1B**) that is consistent with published literature (1, 2, 17, 19). Major populations that can be identified include those expected to be enriched in spermatogonia (Spg), Pre-leptotene spermatocytes (PreL), Leptotene/Zygotene (L/Z) spermatocytes, Pachytene/Diplotene (P/D) spermatocytes, meiosis II spermatocytes (MII) and post-meiotic haploid round spermatids (RS) (**Fig 2-1B**, outlined in red). Importantly, our experimental and cytometric analysis setups yield "tight" demarcation of these spermatogenic populations. This is evident from the formation of distinct L/Z, P/D, MII and RS populations, when viewed using contour plot of one million cells (**Fig 2-1B**, red circles).

To the debris-excluded cells, dead cell exclusion is applied based on PI fluorescence (**Fig 2-1C**). Next, we can additionally exclude unwanted cells based on DNA-content (C-value) by setting boundaries on the Hoechst Blue gate (**Fig 2-1D**). Routinely, we exclude haploid cells with 1C DNA content from our analysis (**Fig 2-1D**). Notably, the prominent delineation between L/Z and P/D in our analysis is marked by the presence of the bimodal 4C DNA content peak on the

Hoechst Blue histogram (**Fig 2-1E**), where left peak corresponds to the L/Z and right peak to the P/D population. Recently, a similar bimodal peak has been documented in guinea pig spermatocytes, whose testis exhibit a higher proportion of L/Z cells, as compared to mouse and rat [102]. Our analysis suggests that despite the underrepresentation of early meiocytes in adult murine testis, consistent dissociation and Hoechst staining conditions, followed by careful discrimination of debris, and noise like elongated spermatozoa, can significantly improve the resolution of the testicular single cell Hoechst profile.

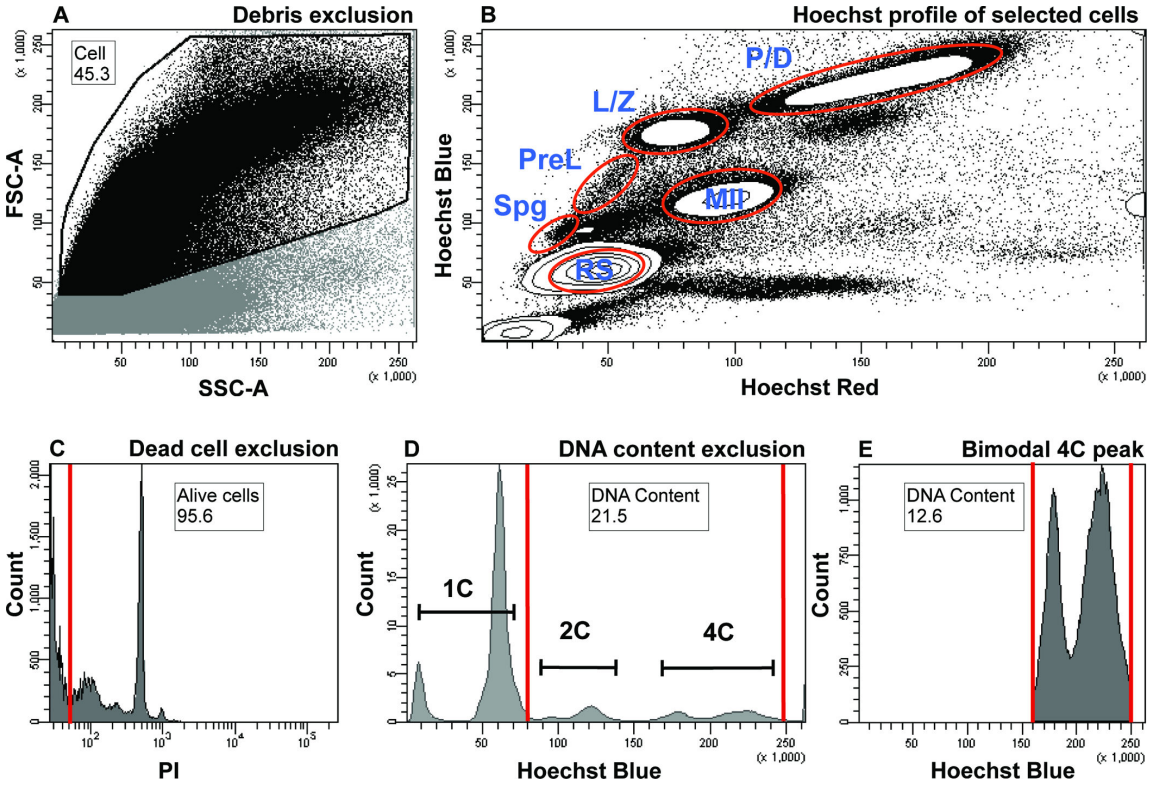


Fig 2-1. Flow cytometric analysis of adult murine testicular cells based on Hoechst and PI fluorescence and light scattering parameters. Numbers on plots represent percent of parent population, the latter identified within a figure legend below. Gate name is found above the number. Both, the number and the gate name are encased by black box. **A)** Debris exclusion based on low light scattering parameters. Cells are distinguished from debris based on the FSC and SSC, proportional to the cell size and cell granularity, respectively. A dot plot shows debris exclusion gate ("Cell" gate, black outline) that includes the cells (black dots) and excludes the debris (grey dots), which exhibit low FSC intensity. The excluded region also contains some elongated spermatozoa, whose small size is a major contributor to the low FSC signal. The parent of the "Cell" gate includes all the cells. **B)** Hoechst profile of testicular cells. Cells selected in Fig 2-2-1A are visualized in a "Hoechst Blue"/"Hoechst Red" contour plot, in which the density of the cells is displayed as contour lines that form circular contours upon high cell density. The main subpopulations visualized are contained within the white densities outlined in red. Spg, spermatogonia; PreL, preleptotene spermatocytes; L/Z, leptotene/zygotene spermatocytes; P/D,

pachytene/ diplotene spermatocytes; MII, meiosis II spermatocytes; RS, round spermatids. **C)** Dead cell exclusion based on PI fluorescence. Alive, PI-negative cells are found within an "Alive cells" gate to the left of the red line and include over 94% of all cells (most are pushed off the x-axis). Cells positive for PI (to the right of the red gate) are excluded from the analysis. The parent of the "Alive cells" gate is the "Cell" gate from Figs 2-1A and 2-1B. **D)** DNA content exclusion based on "Hoechst Blue" fluorescence. Populations that fall within the red gate called "DNA Content" are included in the analysis (2C and 4C DNA contents are labeled). Haploid cells with 1C DNA content are outside of the gate and are excluded from the analysis. The parent of the "DNA Content" gate is the "Alive cells" gate from Fig 2-1C. In this example, the "DNA Content" gate represents 43.3 percent of all cells. **E)** Bimodal distribution of cells with 4C DNA content shows L/Z and P/D populations. The left and right peaks encompassed by the red gate (a restricted "DNA Content" gate) correspond to L/Z and P/D populations, respectively. The parent of the "DNA Content" gate is the "Alive cells" gate from Fig 2-1C. In this example, the restricted "DNA Content" gate represents 5.4 percent of all cells.

Gating and back-gating strategies for isolating high-purity MPI populations

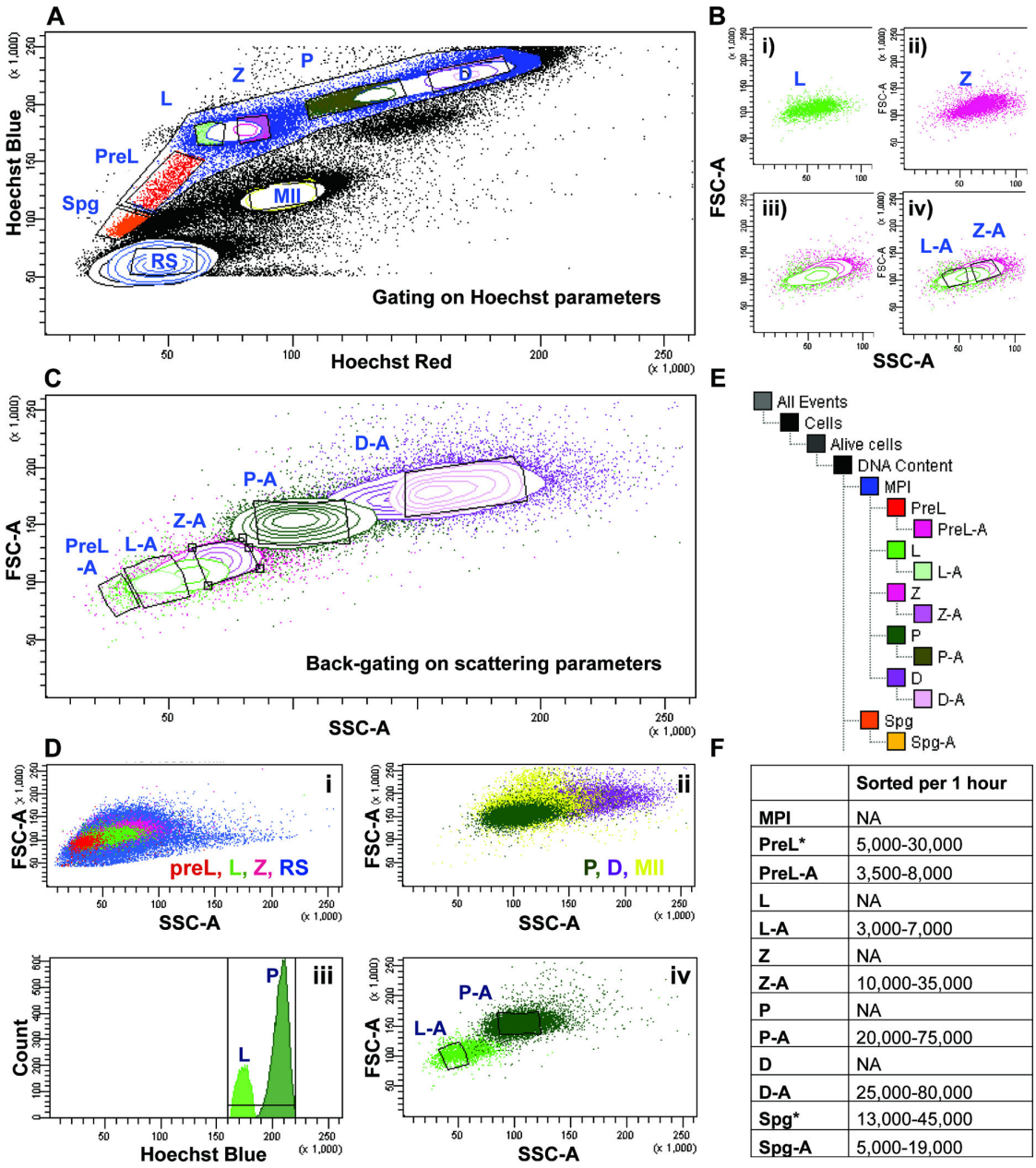
Previously published sorting approaches of murine MPI populations were based on gates specified on Hoechst dye fluorescence, similar to the gating approach in Fig 2-1A–C. This selection alone, however, does not discriminate sufficiently to obtain high-purity L, Z, P or D cells, and, in our experience, these cell populations were found contaminated with each other and with other spermiogenic cell types like haploid spermatozoa. To specify each MPI population more accurately, we wanted to better define the flow cytometric selection gates to minimize contamination. To this end, we have developed an analysis workflow of Hoechst dye-labeled testicular cell suspension that utilizes a back-gating approach to enrich for high-purity MPI substages. The back-gating approach involves setting a gate around the cells with certain fluorescence characteristics, and then asking what are the physical characteristics of these selected cells, based on their FSC and SSC criteria [103]. Ultimately, this approach allows to define a particular cell population of interest not only in terms of two Hoechst fluorescence parameters (Hoechst Blue and Hoechst Red), but also, in terms of the light scattering parameters (FSC and SSC), allowing a more precise definition of that population.

In our standard workflow, we begin by setting the gates on Hoechst dye fluorescence profile (**Fig 2-2A**). Information on gate statistics for Fig 2-2 can be found in Supplementary Fig 2-3. Within a large MPI gate, we assign PreL, L, Z, P and D gates based on previously published and empirical observations. We also gate on non-meiotic populations, including pre-meiotic diploid spermatogonia (gate Spg), diploid spermatocytes II (gate MII), and haploid, post-meiotic spermatids (gate RS) (**Fig 2-2A**). Sorting based on the debris and dead cell exclusion, and on Hoechst dye fluorescence, led to an extensive compromise in purity. Extensive cross-contamination from P was observed in most fractions examined, with Z and D fractions often containing more than 30% of P spermatocytes. Additionally, Spg, PreL and L fractions were found heavily cross-contaminated. Furthermore, most fractions contained some elongated spermatozoa. The overall contamination is largely explained by the proximity and overlap of the

contaminating populations based on FSC and SSC parameters, or the overlap in cell size and granularity. To best illustrate this point, we can examine L and Z populations (**Fig 2-2B**). When L cells from Hoechst dye profile (**Fig 2-2A**, gate L, green) are viewed on the light scatter plot (**Fig 2-2B**, panel i, green), and the same is done with the Z population (**Fig 2-2A and 2B**, panel ii, pink), together, the profiles exhibit a large region of overlap (**Fig 2-2B**, panel iii). Additional gates on the FSC versus SSC profile can be set to more conservatively describe L or Z populations, namely, gates "L-A" and "Z- A", respectively (**Fig 2-2B**, panel iv). These back-gates minimize cross-contamination based on light scattering parameters, and also, incorporate discrimination based on Hoechst dye parameters (**Fig 2-2B**, panel iv). Thus, in this back-gating approach, a gate set on "Hoechst Blue"/"Hoechst Red" fluorescence plot is used to set a gate based on the "FSC"/"SSC" plot. This approach effectively discriminates cells based on both Hoechst dye fluorescence and light scatter parameters.

As MPI substages represent successive stages of germ cell development, there is an extensive overlap on FSC versus SSC plot, between all adjacent MPI populations, not only L and Z. Consequently, we applied the back-gating protocol to all MPI populations (**Fig 2-2C**). In addition, on the "FSC"/"SSC" plot, RS population (**Fig 2-2D**, panel i, blue), which is abundant and morphologically diverse, overlaps with PreL, L and Z populations (**Fig 2-2D**, panel i). Similarly, MII cells (**Fig 2-2D**, panel ii, yellow), overlap with P and D populations (**Fig 2-2D**, panel ii). To accommodate for the overlaps from RS and MII cell types, we additionally discriminate at the level of DNA content, on the "Hoechst Blue" histogram (**Fig 2-2D**, panel iii). Thus, our typical two-way sort, e.g. of L and P populations, will involve a) Debris and dead cell exclusion, b) Exclusion based on DNA content, where the gate is limited to 4C on "Hoechst Blue" histogram (**Fig 2-2D**, panel iii) and c) setting of the back- gates "L-A" and "P-A" on the "FSC"/"SSC" dot plot to specify the final L and P populations before sorting (**Fig 2-2D**, panel iv). Our final sorting tree is depicted in Fig 2-3E. The range of cell numbers collected in an hour, can be found in Fig 2-2F, and contain the numbers from up to ten successful sorts, using mice of

various ages. Typically, we couple the sorting of PreL or L with P, and Z with D. Indeed, some populations like P and D can be collected much faster, without much sacrifice to purity, if they are sorted without the rate- limiting L or Z cells. Also note, that Spg and PreL populations can be sorted without back- gating (**Fig 2-2F**, as indicated by "*"), if sorted alone and with strictly limited DNA content gate, in which case, the numbers collected would also rise.



	Sorted per 1 hour
MPI	NA
PreL*	5,000-30,000
PreL-A	3,500-8,000
L	NA
L-A	3,000-7,000
Z	NA
Z-A	10,000-35,000
P	NA
P-A	20,000-75,000
D	NA
D-A	25,000-80,000
Spg*	13,000-45,000
Spg-A	5,000-19,000

Fig 2-2. Gating and back-gating strategies for isolating MPI populations. **(A)** Gating on individual spermatogenic populations based on Hoechst fluorescence. A large meiotic gate encompasses smaller gates containing cells of individual MPI substages, including preleptotene- (PreL, red), leptotene- (L, green), zygotene- (Z, pink), pachytene- (P, dark green) and diplotene- (D, magenta) spermatocytes. Gates enriched in pre-meiotic spermatogonia (Spg, orange), round spermatids (rSP, blue) and meiosis II spermatocytes (MII, yellow) are also outlined. **(B)** Back-

gating approach. A particular gate on the Hoechst fluorescence plot (Fig 2-3A) can be viewed on the FSC vs. SSC plot. Here, L spermatocytes (panel i) and Z spermatocytes (panel ii) defined by a fluorescence gate in Fig 2-3A, display particular characteristics on the "FSC"/"SSC" plot (green and pink dots, respectively). iii) When viewed on the same plot, L and Z share similar light scattering parameters and partly overlap. iv) Based on regions of minimal overlap on the "FSC"/"SSC" plot, "L-A" gate is created to restrict contamination from the Z gate, and "Z-A" gate is made to restrict contamination from the L gate. **(C)** Back-gating approach applied to all MPI substages. Individual spermatogenic populations defined by gates on Hoechst fluorescence plot ("Hoechst Blue"/"Hoechst Red", Fig 2-3A) are used to set gates on the light scattering plot ("FSC"/"SSC"). A gate set on "FSC"/"SSC" plot and appended with "-A" (e.g., P-A) marks a "back-gate" of a gate (e.g. "P") set on the "Hoechst Blue"/"Hoechst Red" plot. **(D)** DNA content-restricting gate helps eliminate contamination from non-MPI cell types. The light scattering profile of MPI spermatocytes overlaps with that of other spermatogenic cells in the testis. (i) Abundant and morphologically diverse haploid rSP cells (blue) overlap with PreL, L and Z cells. (ii) MII spermatocytes (yellow) overlap with P and D cells. (iii) A gate specifying DNA content on "Hoechst Blue" histogram can be restricted to include only the 4C content where L and P cells are found, and to exclude haploid and diploid cells that fall outside of this gate. Thus, for example, a two-way sort for L and P involves exclusion based on "Hoechst Blue" parameters and on (iv) specification of back-gates "L-A" (light green, outlined subset) and "P-A" (dark green, outlined subset). **(E)** Gating tree. The tree indicates the sequential gating and back-gating procedure applied to Hoechst-labeled testicular single cell suspension before sorting. Gates that have "-A" appended to them were the final sorting gates, such that sorting of L spermatocytes involved collecting cells from the "L-A" gate. **(F)** The range of numbers of cells collected from ten different sorts. The wide range largely reflects the difference in the ages of mice used (2–5 months old) and the adjustment, from experiment to experiment, in the size of gates and back-gates set. Populations that can be sorted without back-gating are indicated by "*".

Detailed characterization and purity assessment of sorted MPI cells

To assess the identity and the extent of homogeneity of sorted MPI cells, we performed co-immunostaining of sorted cells with well-characterized nuclear markers of MPI. Specifically, we examined the distribution of phosphorylated histone H2AX (γH2AX) and synaptonemal complex protein 3 (SYCP3), markers of double stranded-breaks (DSBs) and meiosis-specific synaptonemal complex (SC), respectively. We also used DNA stain DAPI to visualize these cells' dynamic MPI chromatin distribution. Representative MPI substages are shown in Fig 2-3, as imaged by the confocal fluorescence microscopy. Profound changes in chromatin dynamics are marked by the changing shape and number of DAPI-labeled chromocenters (**Fig 2-3**). The assembly of axial elements to which sister chromatids are attached, and the SC, which connects homologous chromosomes, are marked by progressive elongation and thickening of SYCP3. Indeed, SYCP3 aggregates are observed in PreL spermatocytes and long thick SYCP3 fibers in P spermatocytes, the latter indicative of fully synapsed autosomes with assembled SCs (**Fig 2-3**). Increasing amounts and intensity of γH2AX signal in early meiocytes, namely, the PL, L and Z, mark DSB initiation and accumulation, while disappearance of γH2AX in late MPI reflects DSBs resolution, with an exception of an intense, well defined γH2AX signal in P spermatocytes corresponding to partially synapsed sex chromosomes within the sex body (**Fig 2-3**).

Classification of sorted cells, based on the intensity and distribution of DAPI, γH2AX and SYCP3 signals, can be straightforward when the observed cells exhibit known distribution of these markers. Indeed, many of these marker patterns have been previously observed [27, 104-106]. However, examination of hundreds of cells after sorting reveals a number of "non-typical" cells, commonly observed in our sorts, whose MPI substage classification can be challenging. These cells, which have not been well documented in literature, include cells that transition from one MPI substage to another and those at the very beginning or the end of a substage. Thus, the "non-typical" category may include early and late PreL cells, early and late L cells and late L cells transitioning to Z.

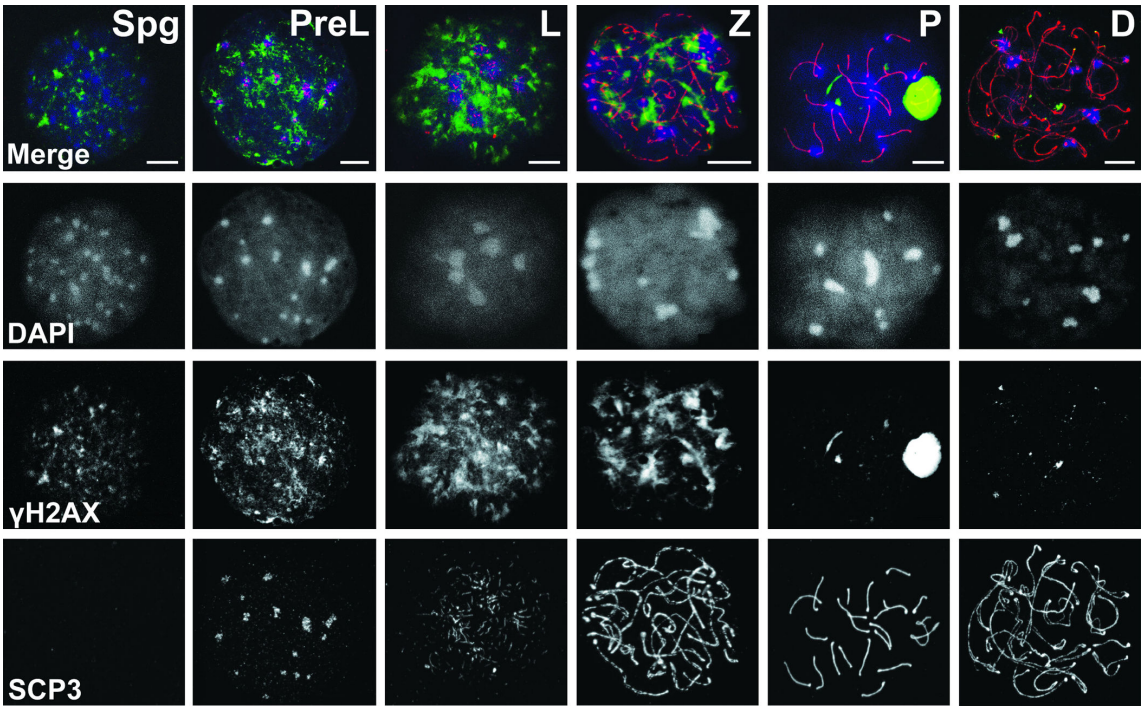


Fig 2-3. Immunofluorescence analysis of MPI progression. Spread nuclei were double labeled with γH2AX (green) and SYCP3 (red) and counterstained with DAPI (blue). Staging was deduced from changing, stage-specific labeling patterns of these markers. Fluorescence images generated by confocal microscopy show representative MPI substages including the preleptotene- (PreL), leptotene- (L), zygotene- (Z), pachytene- (P), diplotene- (D) spermatocytes, and an example of pre-meiotic Spermatogonia (Spg). Bar - 10 μm.

Routine and extensive post-flow cytometry IF analysis of cells based on DAPI staining, SYCP3 and γH2AX, allowed us to better document the diversity of MPI cells. For such routine examination, we used upright fluorescent microsope, since, confocal microscopy, while producing images of very high quality and resolution, is an involved technique that precludes quick cell-by-cell analysis of tens or hundreds of sorted cells, as is required after each sort. Based on IF staining, we defined approximate early-, mid- and late- boundaries of early MPI substages, as they appeared in our partly squashed cell preparations of meiotic nuclei (**Fig 2-4**). Indeed, early through late PreL, L and Z spermatocytes exhibit unique patterns of DAPI, γH2AX and SYCP3 staining (**Fig 2-4** and Figure Legend) and a combination of these patterns can be used as a staple of a particular substage. As there is no ambiguity related to the identification of P and D spermatocytes, we do not present their characterization here.

The categorization based on IF is both, consistent with, and elaborates on previous descriptions of MPI substage specifications, with only a few examples of the latter shown in Fig 2-4. For instance, we observe that late PreL (lPreL) cells often exhibit diffuse and irregular SYCP3 patterns. We label these as late PreL cells because they most often appear in our PreL-specific sorted fraction. However, these lPreL cells can also be considered as transitional PreL to L cells, since they exhibit few to numerous SYCP3 aggregates (**Fig 2-4C**). Consistently, similar cells, with SYCP3 aggregates have been previously placed in either PreL [84, 104-107] or L [84, 104-108] category. Some ambiguities associated with substage identification still remain. One is the difficulty to distinguish with certainty SCP3-negative and γH2AX-negative early PreL cells from differentiated B Spg, which give rise to the PreL cells.

Extensive categorization based on the described markers was critical for subsequent quantification of percent purity of cells after sorting (**Table 2-1**). Significantly, most spermatogenic populations sorted by our approach are of very high-purity, with a range of 75 to 95 percent (for individual data and statistical analysis see Supplementary Fig 2-4). In all the sorted fractions, we were able to greatly reduce contamination from non-meiotic and post-meiotic

cell types. Importantly, we were able to enrich for individual L and Z spermatocytes, enabling separation of now all individual MPI spermatocytes in high-purity form.

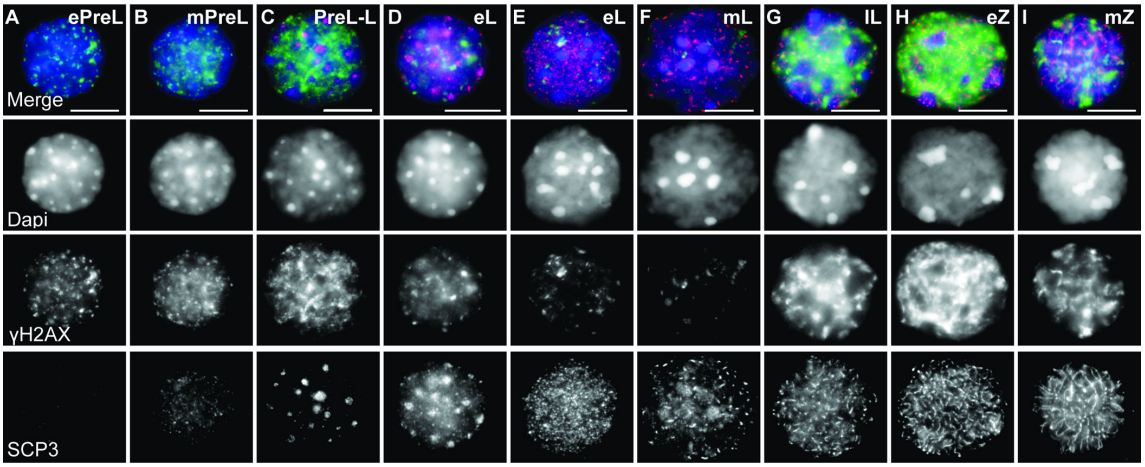


Fig 2-4. Immunofluorescence characterization of early MPI sorted cells. Images show examples of sorted PreL- (A–C, preleptotene), L- (D–G, leptotene) and Z- (H– I, zygotene) spermatocytes. Staging was deduced from DAPI, SYCP3 and γH2AX patterns. Early to mid PreL cells exhibit numerous peripheral DAPI chromocenters corresponding to satellite DNA (A–C). **(A)** Early PreL (ePreL) nucleus shows punctate γH2AX foci and absence of SYCP3. **(B)** Mid PreL (mPreL) nucleus showing foci and patches of γH2AX and weak, diffuse SCP3 staining and **(C)** a nucleus with late PreL and L characteristics (PreL-L) exhibits numerous SYCP3 aggregates and a decrease in DAPI foci along the rim of the nuclear periphery. PreL-L cells often exhibit intense γH2AX signal. **(D–G)** From early L (eL) to late L (lL) nuclei exhibit a progression from short to longer stretches of SYCP3 and from sparse foci to large intense and partly homogenous γH2AX. At least two types of eL cells can be observed, one PreL-like (D) but with lower γH2AX signal, and a more typical one (E). **(H,I)** In early Z (eZ), long, interrupted SYCP3 fibers are observed throughout the cell. Polarized concentration of thickening SYCP3 fiber ends marks telomere bouquet base. Mid Z (mZ) exhibits long, thin SYCP3 stretches. By late Z, chromosome axes are formed and appear as long thin fibers; levels of γH2AX decrease. Bar - 10 μm.

COLLECTED / OBSERVED	Spg	PreL	L	Z	P	D
Spermatogonia (Spg)	80-91%	5-10%	3%	0	0	0
Preleptotene (PreL)	5-10%	75-92%	5-10%	0	0	0
Leptotene (L)	0	5-10%	60%-80%	5-10%	0	0
Zygotene (Z)	0	0	10-15%	75-90%	3%	2%
Pachytene (P)	0	0	3%	10-15%	81-95%	4-10%
Diplotene (D)	0	0	0	0	5-7%	82-95%
Other	4-5%	1-2%	1%	1%	1-2%	2-3%

Table 2-1. Percent purity quantification based on immunofluorescence analysis after cell sorting. MPI substage purity was calculated as (cell type observed/total cells counted by IF)*100. MPI substage purity was assessed based on at least eight different sorts, with a range of 30-150 cells counted per sorted population. The smaller counts are representative of very pure populations, where minimal cell counts were required to establish purity.

Conclusion

In summary, we provide an optimized Hoechst-based flow cytometry method for isolating individual MPI populations from Hoechst dye labeled testicular cell suspension. We offer improvements in sample preparation and show that back-gating technique and DNA content restriction can significantly improve the resolution and separation of early MPI meiocytes. It should be noted, however, that a user of the protocol will probably have to do several flow cytometry sorting experiments to best adjust the gates and back-gates. To evaluate sorted MPI cell purity, we urge users of flow cytometry methodology to employ immunofluorescence staining routinely, and resolve ambiguities with the markers mentioned above, and other meiotic and spermatogonial markers that are available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material (doi:10.1002/cyto.a.22463).

Acknowledgments

This work was performed with the support from Carnegie Institution for Science. We thank Dr. Hao Zhang (at the Johns Hopkins Bloomberg School of Public Health) for the help in early stages of the project and Dr. Phillipe Bois for personal communication and advise on flow cytometry.

Chapter 2.2 Protocol for flow cytometry of murine spermatocytes

Valeriya Gaysinskaya[1,2] and Alex Bortvin[1]*

[1] Department of Embryology, Carnegie Institution for Science, Baltimore MD 21218.

[2] Department of Biology, Johns Hopkins University, Baltimore MD 21218

*Correspondence: Department of Embryology, Carnegie Institution for Science, 3520 San Martin Drive, Baltimore MD, 21218

Phone: (410) 246-3034

Fax: (410) 243-6311

E-mail: bortvin@ciwemb.edu

Adapted and partially reprinted from *Current Protocols in Cytometry* 7.44.1-7.44.23, April 2015, entitled "Flow Cytometry of Murine Spermatocytes".

Published online April 2015 in Wiley Online Library (wileyonlinelibrary.com).

doi: 10.1002/0471142956.cy0744s72 .

ABSTRACT

Protocols for purification of murine male germ cells by FACS based on Hoechst 33342 (Ho342) dye staining have been reported and optimized. However, the protocols are often challenging to follow, partly due to difficulties related to sample preparation, instrument parameters, data display, and selection strategies. In addition, troubleshooting of flow cytometry experiments usually requires some fluency in technical principles and instrument specifications and settings. This unit describes setup and procedures for analysis and sorting of male meiotic prophase I (MPI) cells and other germ cells. Included are procedures that guide data acquisition, display, gating, and back-gating critical for optimal data visualization and cell sorting. Additionally, a flow cytometry analysis of spermatogenesis-defective testis is provided to illustrate the applicability of the technique to the characterization and purification of cells from mutant testis.

INTRODUCTION

A flow cytometer measures and reports on multiple characteristics of illuminated single cells as they flow in a fluid stream. Based on characteristics such as cell size, relative complexity, and relative fluorescence intensity, the cells can be enriched and sorted. Details on these and many other aspects of flow cytometry are described elsewhere (e.g., Givan, 2011). Practically, however, a user of FACS technology must be concerned with challenges related to flow cytometry experiment setup, analysis, and troubleshooting.

FACS offers a unique opportunity to study cells of an organ as complex as the adult testis. Indeed, FACS is the dominant technology for purification and enrichment of particular male germ cell types. A major aspect of testis complexity is cellular heterogeneity in terms of cell type, size, morphology, DNA content, and developmental stage. Because of this inherent cell complexity as well as other aspects of testis biology, including the simultaneous presence of different stages of germ cell development and an unequal abundance of cell types, FACS from an adult testicular cell suspension requires strategic enrichment for a particular population of interest. The vital

nucleic acid−binding dye Hoechst 33342 (Ho342) enables the enrichment of individual mouse meiotic prophase I (MPI) populations using FACS [28, 79, 81, 82]. This takes advantage of the dye's unique properties, including wide fluorescence emission spectra, sensitivity to chromatin state and integrity, and fluorescence shifts that are associated with differences in Ho342 concentration within the cell [109, 110]. Furthermore, the accessibility of DNA to various fluorochromes varies at different stages of spermatogenesis, due to changes in chromatin structure during differentiation [111]. Indeed, the differences in chromatin compaction among cells in MPI and between other testicular cell types makes them amenable to resolution by FACS after staining with Ho342. This unit presents hands-on aspects of flow cytometry analysis and cell sorting of a Ho342-stained testicular cell suspension.

Basic Protocol 1 describes the optimization of photomultiplier tube (PMT) voltages in a sample-specific manner using controls. This first procedure in a flow cytometry protocol is important for placing Ho342/PI-stained testicular cell suspension on appropriate scales for forward scatter (FSC), side scatter (SSC), Ho342-blue, Ho342-red, and PI detection, so as not to miss either the small and dim cells or the big and bright cells during analysis. Basic Protocol 2 describes how, using these optimized PMT voltages, to achieve a working flow cytometry profile of a Ho342/PI-stained testicular cell suspension using a series of basic gates drawn around cells of interest. This profile is a founder profile for other more selective and specific gating strategies. Basic Protocol 3 develops a gating strategy to specify testicular subpopulations of interest. This protocol utilizes gates and back-gates to analyze and subsequently sort individual MPI subpopulations (including pre-leptotene [PL], leptotene [L], zygotene [Z], pachytene [P], and diplotene [D] spermatocytes), as well as other testicular populations (including premeiotic spermatogonia [Spg]). The strategy is optimized to yield significant enrichment of these populations. Basic Protocol 4 describes Ho342/PI-based flow cytometry analysis of mutant testis to demonstrate the applicability of the setup and gating practices established on wild-type (WT) testis. Support Protocol 1 describes preparation of testicular cell suspensions for FACS, and

processing for staining and flow cytometry. Finally, Support Protocol 2 describes how analyze the purity of sorted cells using immunofluorescence with well-characterized germ cell markers.

BASIC PROTOCOL 1: Optimization of PMT voltage using control cells

The photodetection system of a cytometer consists of PMTs along with wavelength-specific mirrors and filters. The mirrors and filters partition the light emitted from cells onto a series of separate PMTs according to wavelength, so that the PMTs detect signals from different wavelengths. PMTs detect and amplify weak signals generated by scatter and fluorescence light. The signal amplification is achieved by applying a voltage to the PMTs (amplification gain). Baseline PMT voltages are set as part of the cytometer performance, which is measured daily according to the manufacturer's instrument-specific guidelines. BD Biosciences Software, used in this protocol, utilizes a Cytometer Set- tings and Tracking (CS&T) assay to set color-specific PMT values for the detection of a dynamic range of fluorescence intensities (from dim to bright). Since baseline PMTs are typically optimized using fluorescent beads rather than cells, it is important to also optimize PMT settings for the cells of interest in order to establish the appropriate fluorescence detection range. Optimizing PMT voltages for a specific experimental setup is therefore critical for optimal fluorescence signal detection, sensitivity, and resolution.

The objective of this protocol is to optimize PMT voltages for cells stained with Ho342 and propidium iodide (PI) in order to establish the detection range of negatively and positively stained cells, and to minimize the contribution of electronic noise to the signal. Specifically, the user is instructed to set optimal FSC and SSC voltages and fluorescence PMT voltages using four control samples of murine testicular cell suspensions: unstained, single stained (PI or Ho342), and double stained (Ho342 and PI). These controls are also used to help define positive/negative boundaries in the data, providing controls for alive and dead cells, and stained and unstained cells. These boundaries can be used to eliminate unwanted populations from the start. The optimized control settings and gates can be reused from experiment to experiment, provided that the

cytometer settings remain the same.

Fluorochrome Specifications

Fluorochrome	Excitation/Emission max (nm)	Comment
Hoechst 33342 (Ho342)	350/461 max (bound to DNA)	* Wide emission spectrum allows Ho342 Red detection
Propidium Iodide	540/607 (bound to DNA)	

Laser and Optics Device Configuration

Laser and optics device	PMTs	Longpass mirrors	Bandpass filters	Fluorochromes
Blue 488-nm Trigon	A	655LP	695/40	N/A here; otherwise, PerCP-Cy5.5, PerCP
	B	502LP	530/30	N/A here; otherwise, FITC
	C	—	488/10	SSC
Red 633-nm Trigon	A	735LP	780/60	N/A here; otherwise, APC-Cy7, APC-H7
	B	—	660/20	N/A here; otherwise, APC
Near UV 375-nm Octagon	A	610LP	670LP	Ho342 Red, propidium Iodide
	B	—	450/40	Ho342 Blue; otherwise, DAPI
	C	502LP	510/50	N/A here; otherwise, AmCyan

Table 2-2. Fluorochrome specifications (top) and laser and optics device configuration (bottom).

Materials

Control samples (see Support Protocol 1)
Unstained cells, Ho342-stained cells, PI-stained cells, Ho342/PI-stained cells

FACS cytometer (BD FACSAria III or similar) equipped with:
Excitation optics: near-UV (375-nm) and blue (488-nm) lasers
Detection optics: filter combination consisting of 450/40 bandpass
(Ho342-blue detection), 670 longpass (Ho342-red and PI detection), and 610 dichroic longpass mirror (beam splitter)

FACS analysis software (BD FACSDiva or similar)

*See **Table 2-2** for fluorochrome specifications and laser and optics device configuration*

1. Create the following working plots displaying parameters to be optimized:

FSC-A vs. SSC-A plot on a linear scale (dot plot or a contour plot display)

PI histogram plot on a logarithmic scale

Ho342-blue vs. Ho342-red on a linear scale (contour plot display)

In the sample study shown in Fig 2-5, FSC, SSC, and Ho342 fluorescence signals were acquired with linear amplification and plotted on a linear scale. The data plotted show the area measurements (specifically, the integrated intensity) of signals, with the FSC and SSC axis labels appended with "-A" to indicate area. PI fluorescence signals were also acquired with linear amplification but plotted on a log scale. Whether or not the acquired signal amplification is linear or logarithmic is a property of the cytometer and refers to the way the signals from cells are processed or "summarized" from signals to data.

See Critical Parameters and Troubleshooting section on data display for information on dot versus contour plots and linear versus exponential or bi-exponential scales.

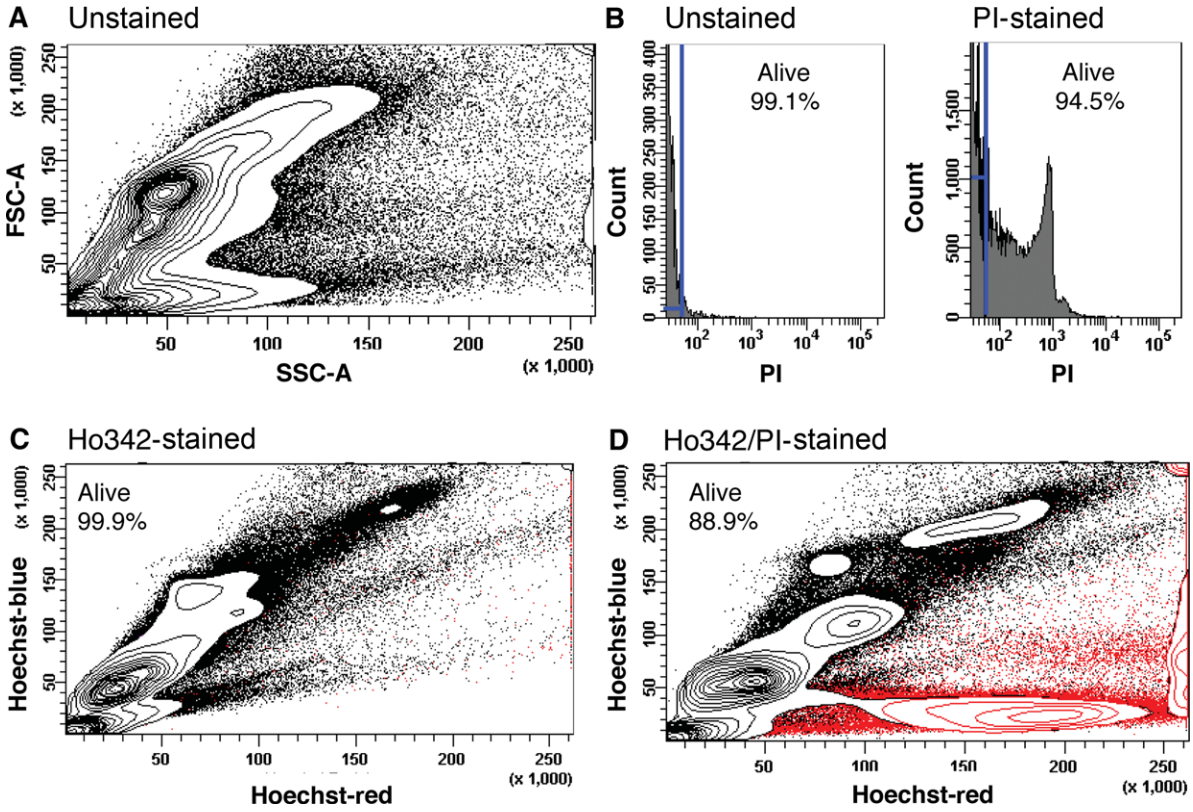


Fig 2-5. FACS analysis of control adult murine testicular cells. All signals were acquired with linear amplification. Signals are plotted as area integrated intensity values. Percent values on plots represent percent of parent population (the latter identified below). Gate name is specified above the percent value. (A) Acquired unstained cells (1×10^6) depicted as a contour (probability) plot of FSC vs. SSC. Cells are bound by contour and are located outside of contour as single dots. These outlier dots usually represent 2.5% to 5% of the total population, depending on the probability of the contour set. (B) Acquired unstained (left, 1×10^6) and PI-stained (right, 1×10^6) cells depicted on a PI histogram plot. Left panel: Live PI-negative cells are found within an "Alive" interval gate (in blue) to the left of the gate. Dead cell exclusion is based on PI fluorescence. Right panel: PI-stained control sample. PI-positive cells (to the right of blue interval gate) are excluded from the gate, and thus from subsequent analysis. The parent of the "Alive" gate includes all cells. (C) Hoechst fluorescence profile of Hoechst-stained cells visualized on a

Hoechst-blue/Hoechst-red contour plot. Since no PI was added, all cells (or 99.9%) appear as live cells. (D) Hoechst fluorescence profile of Hoechst/PI- stained cells visualized on a Hoechst-blue/Hoechst-red contour plot. Cells in red with low Hoechst-blue and mid to high Hoechst-red indicate PI-positive dead cells based on “Alive” gate analysis.

2. Run and acquire the unstained sample on the FSC/SSC plot (**Fig 2-5A**). While the sample is running, optimize FSC and SSC voltages by adjusting both voltages up or down to place the unstained cells on the scale of the FSC/SSC plot.

Adult testicular cell suspensions are complex and diverse in terms of cell size and morphology. The complexity is reflected in the broad range of FSC and SSC values, which are proportional to cell size and complexity, respectively. The aim is to distribute cells broadly on the FSC axis in order to visually maximize the resolution of both small and large MPI cells. If the goal was, instead, to optimize the view of D-spermatocytes (some of the largest cells found in the testis) the user might decrease the FSC voltage. While compacting the smaller cells into a tighter distribution on the FSC/SSC plot, the downward shift in FSC would encapsulate all large cells and expand their resolution on the plot, forcing the overall cellular profile to shift towards the lower left corner of the plot.

3. Run and acquire the unstained sample on the PI histogram plot (**Fig 2-5B**, left panel). Adjust the PI voltage up and down while the sample is running to position the PI-negative population in the first decade of the logarithmic plot.

The PI-negative (alive) population should appear close to zero on the axis.

4. On the PI histogram plot, draw an interval gate called "Alive" around the major PI-negative live cell population (**Fig 2-5B**, left panel, blue gate).

The major histogram corresponding to live cell peak should appear pushed of the axis and bounded by the gate on the right and the y axis on the left. On the four-decade logarithmic axis, the left end of the gate is pushed of the axis to accommodate the rest of the live, PI-negative cells.

5. Run and acquire the PI-stained sample on the PI histogram plot (**Fig 2-5B**, right panel, blue gate). In the presence of dead cells, observe a histogram shift to the right of the "Alive" gate, accompanied by a decrease in the percentage of live cells as measured by the "Alive" gate.

If necessary to better resolve the shifted PI-positive population, adjust PI voltage to move the population to the right (towards a higher decade on the logarithmic scale) and readjust the "Alive" gate boundary.

6. Run and acquire the Ho342-stained sample on the Ho342-blue/Ho342-red plot (**Fig 2-5C**). To optimize the Ho342 fluorescence PMT voltages, adjust Ho342-blue and Ho342-red voltages up and down while running the sample to maximize the resolution of Ho342-blue/Ho342-red cell profile on the linear axis. The population as a whole should appear to tilt diagonally across the

plot.

Since the "Alive" cell gate represents a subset of all acquired events, the total cells acquired on the Ho342 plot (or any other plot you wish to create) have an "Alive" gate value associated with them. The Ho342-stained sample exhibits an "Alive" gate value of 99.9%, because this control has not been stained with PI, and thus PI-positive cells cannot be detected.

7. Run and acquire the Ho342/PI-stained sample on the Ho342-blue/Ho342-red plot (**Fig 2-5D**).

View live cells by examining cells within the "Alive" gate on a Ho342/PI plot (**Fig 2-5D**, black cells and contour lines).

In Fig 2-5D, note that percent live cells decreases to 88.9%. PI-positive cells (in red) have low Ho342-blue in combination with mid and high Ho342-red characteristics. The cells outside the "Alive" gate (PI-positive, dead cells) can be marked by a different color (red in Fig 2-5D). One way to do this is to create an inverted "Alive" gate, which will represent all cells outside of the "Alive" gate. It is often useful to visualize "dead" cells on the Ho3342/PI plot, even though these cells are eliminated by exclusion from the "Alive" cell gate on the PI histogram plot. It is useful to view these PI-positive cells to assure that the major (diagonally leaning) Ho342 cell population doesn't overlap with the dead cells. Otherwise, it would be difficult to separate the two populations, and the PMTs would need to be adjusted to account for that. It is important to remember that the PI-positive population only represents the percentage of dead cells/debris currently present in the final sample, and does not reflect how many cells in total have died. One reason is the use of DNase I enzyme during sample preparation, which eliminates any leaked DNA from cells with broken membranes. Such dead cells may escape true marking by PI, since they may be found in the debris category of the testicular cell suspension profile.

8. Save the cytometer settings with optimized PMT values and apply these settings when running future FACS experiments with Ho342/PI-stained testicular single-cell suspensions.

Optimized PMT voltage settings can be reused from experiment to experiment, until any aspects of the cytometer configuration are changed (e.g., when new baseline PMT values are set).

BASIC PROTOCOL 2: Setup of working flow cytometry profile

This protocol describes a basic gating strategy to set up a working visual profile of Ho342/PI-stained testicular cells. The gating setup can be used as a foundation for all subsequent gating strategies (see Basic Protocol 3). The specific goal of this protocol is to optimize visualization of a Ho342- and PI-stained testicular cell suspension using debris exclusion based on low light scattering parameters, dead cell exclusion based on PI fluorescence, and DNA content exclusion based on Ho342-blue fluorescence.

Materials

FACS cytometer and analysis software (see Basic Protocol 1)

Ho342/PI-stained testicular cell suspension (see Support Protocol 1)

1. Apply optimized PMT settings (see Basic Protocol 1).

2. Create four working plots:

FSC-A vs. SSC-A on a linear scale (dot plot or contour plot display) ;

PI histogram plot on a logarithmic scale

Ho342-blue histogram plot on a linear scale Ho342-blue/Ho342-red on a linear scale (dot plot or contour plot display)

3. Run and acquire at least 500,000 events (cells). Examine acquired cells in the plots above.

4. On the FSC/SSC plot, draw a gate called "Cells" to exclude cells with low FSC and SSC values (**Fig 2-6A**, blue gate).

The "Cells" gate should appear as a "child" (or subset) of all acquired events. The FSC cutoff eliminates debris and some haploid cell types of relatively small size, such as early spermatids. Because haploid cells are highly abundant, they can easily contaminate other testicular cell populations if not eliminated. Events that exhibit low FSC in combination with mid and high SSC values are also eliminated from the analysis, as the eliminated cells are mostly elongated spermatids. Information on the enrichment of elongated spermatids by FACS can be found in [112].

5. Draw an "Alive" interval gate on the PI histogram plot to include PI-negative alive cells and exclude PI-positive dead cells (**Fig 2-6B**, blue gate), based on the threshold established in Basic Protocol 1.

The "Alive" gate should represent a subset of the "Cells" gate.

6. Draw a "DNA" gate on the Ho342-blue histogram (**Fig 2-6C**, blue gate) to refer to DNA content. Manipulate cell inclusion or exclusion based on the Ho342-blue signal proportional to DNA content. Eliminate low Ho342-blue events from the analysis to eliminate debris and unstained cells.

In Fig 2-6C, the peak at <50,000 relative fluorescence units (RFU) was eliminated. The "DNA"

gate should be manipulated based on sorting needs. If haploid cells are not collected, the gate should be drawn to entirely eliminate the Ho342-blue signal corre- sponding to haploid cells (Fig 2-6C, the tallest peak, from 50,000 to 80,000 RFU).

7. View cells on a Ho342-blue/Ho342-red plot, as they appear after serial gate selection using the

"Cells," "Alive," and "DNA" gates.

The result is the working flow cytometry profile of Ho342/PI-stained cells.

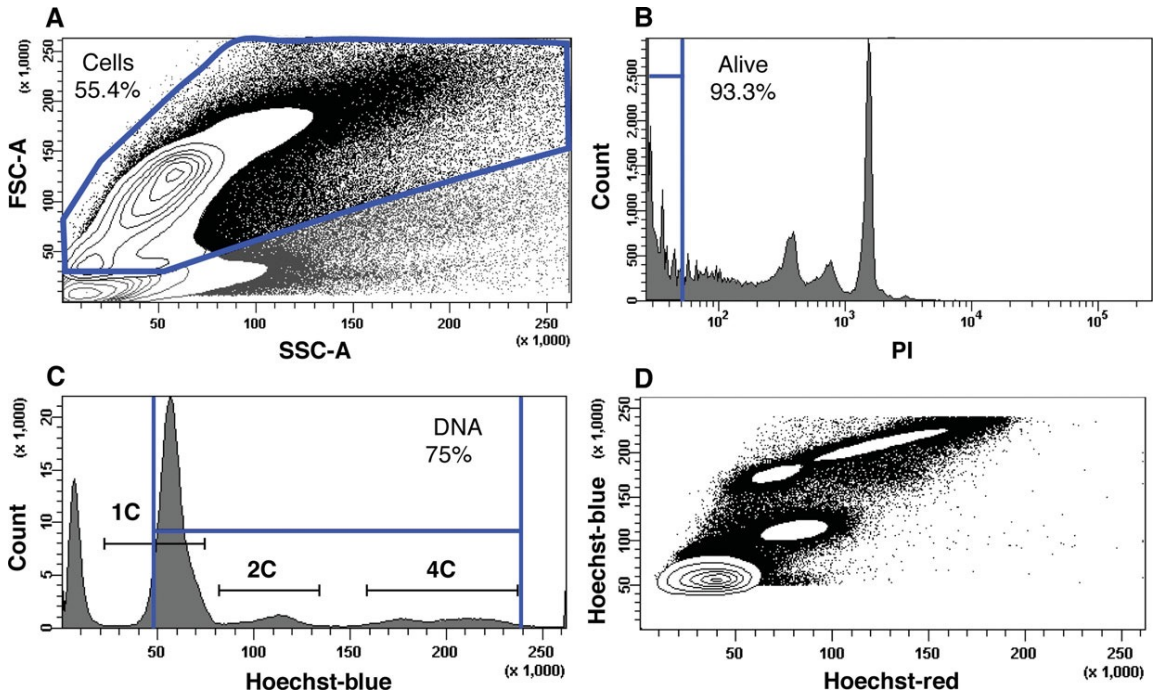


Fig 2-6 FACS analysis of adult murine testicular cells based on Hoechst and PI fluorescence and light scattering. All signals were acquired with linear amplification. Signals are plotted as area integrated intensity values. Percent values on plots represent percent of parent population (the latter identified below). Gate name is specified above the percent value. (**A**) Debris exclusion based on low light scattering. Cells (inside blue gate) are distinguished from debris (outside of blue gate) based on low light scatter viewed on a FSC/SSC contour (probability) plot. "Cell" gate represents a debris- exclusion gate (or cell-inclusion gate) that excludes debris as well as relatively small cells, such as haploid cells. The parent of the "Cells" gate includes all cells acquired (1×10^6). (**B**) Dead cell exclusion based on PI fluorescence. Alive PI-negative cells appear within an "Alive" gate in blue. The parent of the "Alive" gate is the "Cells" gate in (A), thus live cells form a subset of cells selected by the "Cells" gate. (**C**) DNA content exclusion based on Hoechst-blue fluorescence. Populations that fall within the blue "DNA" gate include cells with 1C through 4C DNA content. Low Hoechst-blue events are eliminated and correspond to debris and dead cells. The parent of the "DNA" gate is the "Alive" gate in (B). (**D**) Hoechst

profile of testicular cells after serial selection with “Cells," “Alive,” and “DNA” gates. This working profile of Hoechst/PI-stained testicular cells will be the basis for subsequent gating and back-gating strategies.

BASIC PROTOCOL 3: Setup of gates and sorting of control testicular subpopulations

The gating protocol below is divided into three parts. The first identifies regions of interest based on Ho342 fluorescence. These regions include haploid germ cells (round and elongated spermatids), diploid germ cells (secondary spermatocytes), early MPI cells (L and Z spermatocytes), late MPI cells (P and D spermatocytes), premeiotic diploid Spg, and PL spermatocytes that undergo last replication before meiosis. The second describes gating (on Ho342 fluorescence) and back-gating (on FSC and SSC) strategies for specifying testicular subpopulations based on both Ho342 fluorescence and light scatter, and enriches for individual MPI substages and premeiotic Spg. The third describes sort setup for optimal enrichment and purity of individual fractions. Finally, details are provided for sorting the testicular single-cell suspension using these established gates. General information on principles behind flow sorting, including drop formation and practical aspects of sorting, can be found throughout this book, and elsewhere [103, 113, 114].

Materials

Ho342/PI-stained testicular cell suspension (see Support Protocol 1)
5% (v/v) newborn calf serum (NCS) in Gey's balanced salt solution (GBSS; Sigma), filtered through 40-μm nylon cell strainer
5-ml round-bottom polystyrene tubes 35-μm cell strainer cap (BD Falcon)
FACS cytometer and analysis software (see Basic Protocol 1)

Identify regions of interest based on Ho342

1. Set up a working flow cytometry profile of Ho342/PI-stained testicular cells based on the selection gates drawn on the FSC/SSC plot ("Cell" gate), PI plot ("Alive" gate), and Ho342-blue/Ho342-red plot (see Basic Protocol 2).

Perform gating

2. On the resulting Ho342-blue/Ho342-red plot, draw gates R0 through R7 around regions of interest (**Fig 2-7A**), as follows:

Contour (probability or density) plots are recommended for fluorescence display. On a contour (probability) plot, major populations are located within the contours lines. Con- tour plots commonly show 90% to 95% of data bounded by the contour lines, depending on contour probability set. Contour plots provide more accurate data representation than dot plots, because they can more precisely localize and distinguish subsets of cells [115].

a. Draw the R0 gate around the densest region on the plot, with low Ho342-blue and low Ho342-red characteristics. This region contains haploid sperm. Check that this population maps back to the 1C DNA content on the Ho342-blue histogram plot (**Fig 2-6C**).

*As a rule, gates should be drawn around visually distinct populations. On a contour plot, such populations are bound or separated by contours. The latter case results from populations being not large enough to form a contour-bound density. Similar to the R0 gate, all gates drawn on the Ho342-blue/Ho342-red plot should be viewed on the Ho342-blue histogram plot (**Fig 2-6C**) to ensure that the selected cell population is consistent with the expected DNA content of that population.*

b. Draw the R1 gate around the region with mid Ho342-blue and mid Ho342-red characteristics. On the Ho342 plot, this region is high in cell density and contains secondary spermatocytes with 2C DNA content.

c. Draw gates R2 and R3 by "splitting" the single density contour R2+R3. Cumulatively, this single contour contains early MPI cell types with 4C DNA content (L and Z spermatocytes). Individually, the R2 region (left) contains more L spermatocytes and the R3 region (right) contains more Z spermatocytes.

*The content asymmetry of contour R2+R3 can be reflected in the FSC/SSC plot (**Fig 2-7B**), which portrays profiles of R2+R3 as one gate, and reveals slight density polarization, suggesting a difference in cell scatter characteristics that L and Z spermatocytes are known o exhibit.*

d. Draw gates R4 and R5 by splitting the single density contour R4+R5. Cumulatively, these regions contain late MPI cell types with 4C DNA content (P and D spermatocytes), with R4 (left) containing more P and R5 (right) containing more D spermatocytes.

*As for R2+R3, the content asymmetry of R4+R5 can be observed on the FSC/SSC plot (**Fig 2-7B**) as a single gate with left-right density polarization, suggesting a difference in cell scatter characteristics that P and D spermatocytes are known to exhibit.*

e. Draw R6 gate. This region contains premeiotic diploid Spg with 2C DNA content. In Fig 2-7A, this population "sits" on top of the haploid cell population, such that if you limit the "DNA" gate

on the Ho342-blue histogram (**Fig 2-7C**), eliminating haploid cells from view, and then view the limited cell profile on the Ho342 plot, the R6 region should remain and form a distinct density, as it gains more probability weight once the highly dense haploid density has been eliminated.

f. Draw R7 gate. This region contains PL cells in S phase, with 2C to 4C DNA content.

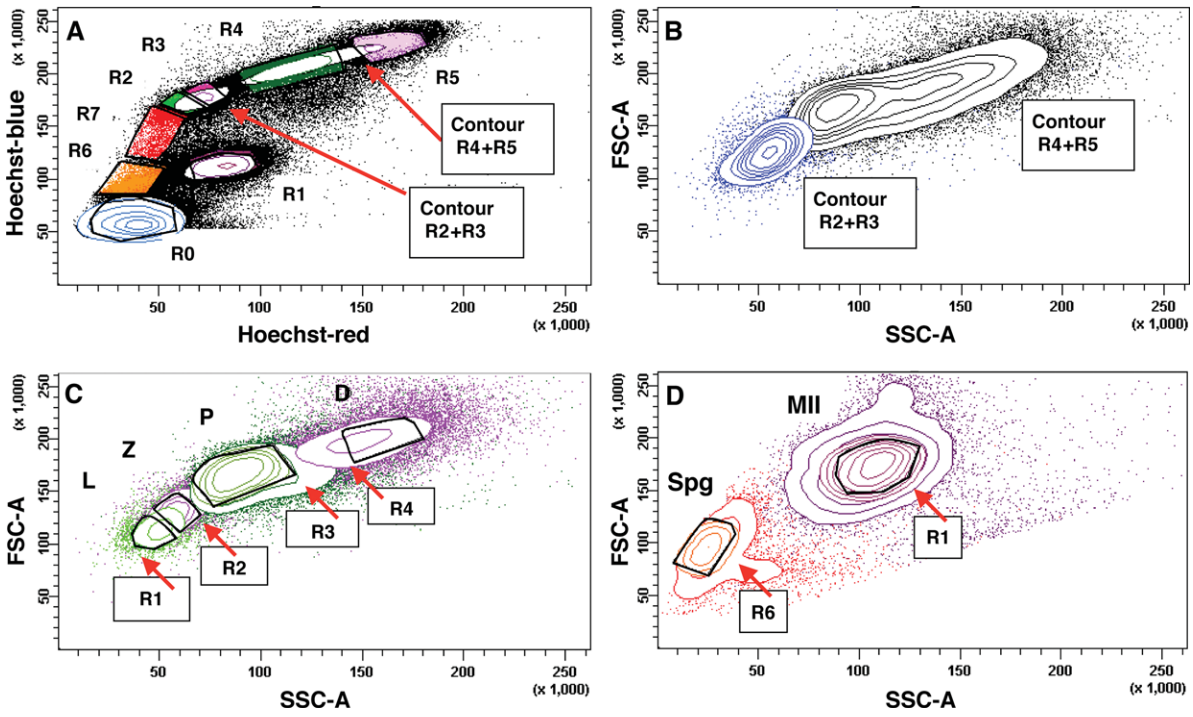


Fig 2-7. Gating and back-gating strategies for isolating individual MPI populations. All signals were acquired with linear amplification. Signals are plotted as area integrated intensity values. (**A**) Gating on individual testicular populations based on Hoechst-blue/Hoechst-red fluorescence. R0 (blue) is enriched in haploid spermatids; R1 (purple) is enriched in meiosis II (MII) spermatocytes; R2 (green) is more highly enriched in leptotene (L) than in zygotene (Z) cells; "R3" (pink) is more highly enriched in Z than in L spermatocytes; R4 (dark green) is enriched in pachytene (P) spermatocytes; R5 (purple) is enriched in diplotene (D) spermatocytes; R6 (orange) contains spermatogonia (Spg); R7 (red) is enriched in pre-leptotene (PL) spermatocytes. The R2+R3 contour represents the region from which R2 and R3 originated and indicates highly similar cells. The R4+R5 contour represents a region from which R4 and R5 gates originated. (**B**) Contours R2+R3 (blue) and R4+R5 (black) from the Hoechst plot (A) are depicted (for analysis only) on the FSC/SSC plot to indicate the polar or lobed nature of these regions, an indication of two potentially distinct populations within a common density. These contours are divided into individual regions (R2, R3, R4, R5) in (A). (**C**) Back-gating involves

visualization of a particular gate from the Hoechst plot (A) on an FSC/SSC plot, where it is possible to observe overlaps between populations based on scatter. Note overlap of R2 (green) with R3 (pink), of R3 with R4 (dark green), and of R4 with R5 (purple). Based on regions of minimal overlap on the FSC/SSC plot, the L, Z, P, and D gates are created to enrich for their respective spermatocyte populations. Back-gates serve to restrict contamination in gates drawn on the Hoechst plot, resulting in more pure populations during sorting. (**D**) Back-gating to enrich for Spg (R6 in A) and MII germ cells (R1 in A). A shift of the R6 population (red) towards increasing SSC values suggests contamination of a more prominent population, which also appears to shift towards increasing FSC values. The smaller contaminating lobe indicates contamination with somatic cells. The Spg back-gate eliminates the contribution of this population from analysis and sorting.

Perform back-gating

3. On the FSC/SSC plot, draw back-gates to further specify the testicular populations of interest.

*Back-gating involves setting gates on regions previously defined based on Ho342 characteristics (R0 through R7) in terms of differences in scatter characteristic. The back-gates drawn on the FSC/SSC plot thus form the "child" gates of the Ho342 gates (**Fig 2-7A**) and can be associated with particular Ho342 fluorescence. This dual selection, first on Ho342 and then on FSC/SSC, minimizes the overlap based on fluorescence and allows for a more narrow and precise definition of particular populations.*

a. Draw back-gates L, Z, P, and D (**Fig 2-7C**).

Note that gates that appear distinct on the Ho342 plot (e.g., R2 and R3, or R4 and R5 in Fig 2-7A) in fact overlap based on their FSC/SSC characteristics (green and purple contour overlaps corresponding to R2/R3, R3/R4, and R4/R5 overlaps in Fig 2-7C). Creation of L, Z, P, and D back-gates separates these populations based on light scattering parameters.

b. Draw back-gates Spg and meiosis II (MII).

In Fig 2-7D, the extension of the original R6 gate (orange) to the right towards increasing SSC, marks contamination with somatic cells. Thus, if the Spg back-gate is not drawn on top of the R1 gate (defined based on Ho342), sorting of the R6 gate would result in a fraction contaminated with soma.

c. Examine the gating tree formed from sequential gating and back-gating (**Fig 2-**8).

Sort testicular subpopulations

4. Perform cytometer-specific preparation for sorting, including optimization of side streams and drop delay [113].

5. Determine populations to sort using a two-way system that allows two populations of interest to be sorted at a time into a collection device that holds two tubes. Refer to gating tree in Fig 2-8 for the following pairwise combinations:

a. L and P spermatocytes based on L and P back-gates

b. L and D spermatocytes based on L and D back-gates

c. PL spermatocytes alone from gate R7

d. PL with P or D spermatocytes

e. Spg spermatocytes from gate R6 alone *or* Spg with MII spermatocytes from back-gate R1

When sorting, limit "DNA content" to contain only the population range of fractions to be sorted. For example, when sorting Z and D populations, limit DNA content from the beginning of Z to the end of D.

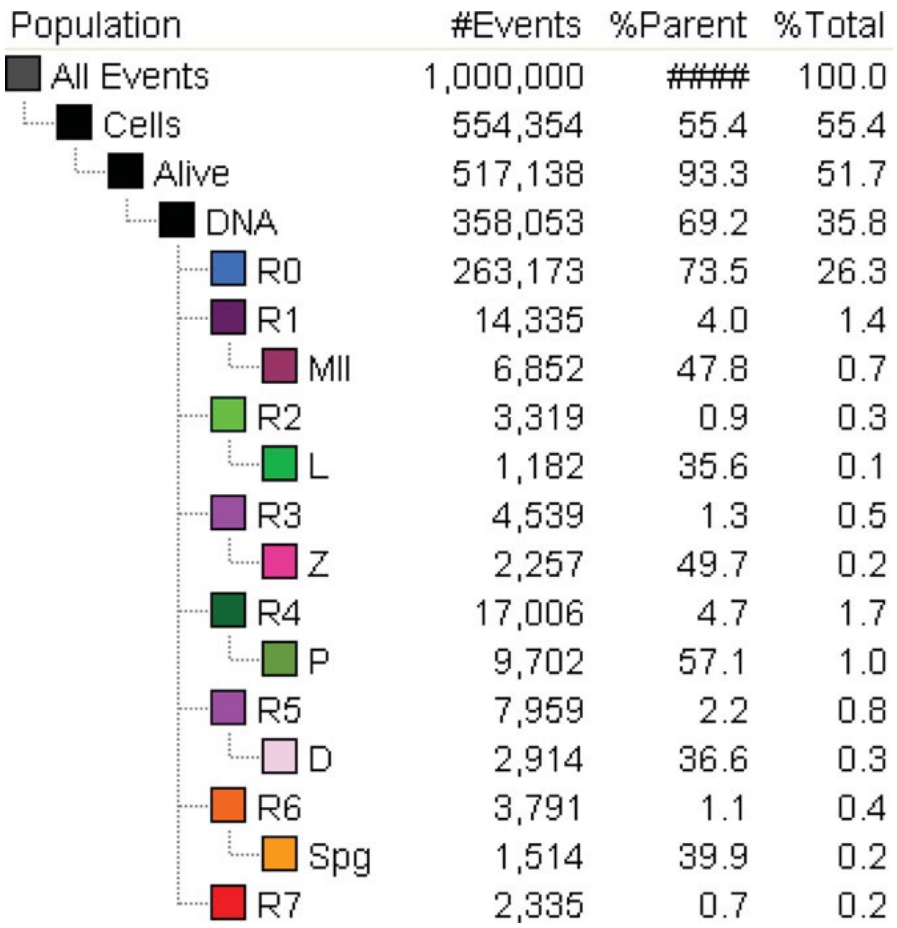
6. Prepare tubes for cell collection during cell sorting. Coat two empty 5-ml polypropylene round-bottom tubes with 5% NCS in GBSS by pipetting with a plastic transfer pipet, leaving 1 ml for cells to drop into.

7. Run the experimental sample (prepared in Support Protocol 1, and used immediately after step 18 of that protocol). Adjust flow rate to 1500 to 2500 events/sec.

The efficiency of sorting is affected by the flow rate, which reflects the speed of sorting. Sorting at high flow rates (>3500 events/sec) results in contamination of sorted cells. When flow rate starts to appreciably decrease with time, and you find yourself needing to increase flow rate, stop to check if the cell suspension has accumulated debris or clumps. Add more DNase I and resuspend, or use a new aliquot of Ho342/PI-stained cell suspension.

8. Acquire and record pre-sort data to adjust sorting gates, if necessary. Pre-record at least 250,000 events. Adjust sorting gates and back-gates if necessary.

9. Proceed with sorting populations of interest based on the sorting gates adjusted in pre-sorting step above.



Population	#Events	%Parent	%Total
All Events	1,000,000	#####	100.0
Cells	554,354	55.4	55.4
Alive	517,138	93.3	51.7
DNA	358,053	69.2	35.8
R0	263,173	73.5	26.3
R1	14,335	4.0	1.4
MII	6,852	47.8	0.7
R2	3,319	0.9	0.3
L	1,182	35.6	0.1
R3	4,539	1.3	0.5
Z	2,257	49.7	0.2
R4	17,006	4.7	1.7
P	9,702	57.1	1.0
R5	7,959	2.2	0.8
D	2,914	36.6	0.3
R6	3,791	1.1	0.4
Spg	1,514	39.9	0.2
R7	2,335	0.7	0.2

Fig 2-8. The gating tree. The gating tree formed from sequential gating and back-gating applied to a Hoechst/PI-stained testicular single-cell suspension before cell sorting.

BASIC PROTOCOL 4: Flow cytometry analysis of spermatogenesis-defective testis

This protocol describes the application of Ho342 and PI staining to flow cytometric analysis of testicular germ cells from spermatogenesis-defective mice. The example here uses a mutant lacking the Maelstrom protein (*Mael−/−*), which is essential for spermatogenesis [116]. Similar analyses can be used to investigate and purify germ cells from other spermatogenesis-defective testes. This analysis demonstrates that the Ho342-stained cell profile of adult *Mael−/−* testis is consistent with the known phenotype, which is characterized by failure to complete MPI and a lack of P and D spermatocytes and all subsequent cells of spermatogenesis [116].

Flow cytometry exploration of mutant populations can be done in two ways: by applying a WT analysis strategy and by performing de novo analysis. In the first scenario, mutant flow cytometry profiles are acquired using WT settings, including PMT values and gates. This analysis is useful for direct comparison of WT and mutant testis profiles. In the second scenario, PMTs and gates are adjusted and set up de novo for the mutant sample. Both scenarios follow steps outlined in the Basic Protocols 1 and 2. However, in the second scenario, PMTs and gates are adjusted independently of WT settings and accommodate mutant testicular cell profile, specifically. In this protocol, the second scenario is explored.

Materials

Ho342/PI-stained *Mael−/−* testicular cell suspension (see Support Protocol 1)

FACS cytometer and analysis software (see Basic Protocol 1)

1. Adjust FSC, SSC, Ho342-blue, Ho342-red, and PI PMT voltages (see Basic Protocol 1).

As in Basic Protocol 1, this step involves acquiring the Ho342/PI-stained sample and adjusting the voltages up/down while running the sample, until the population is on scale.

2. Set up the working flow cytometry profile of Ho342/PI-stained cells, drawing serial gates “Cells," “Alive,” and “DNA Content” as described (see Basic Protocol 2).

*The "Cells" and "DNA Content" gates for Mael−/− cells are shown in blue in Fig 2-9 (panels A and B, respectively). On the FSC/SSC plot (**Fig 2-9A**), the occupied range of FSC and SSC values is smaller for Mael−/− than for the WT testis (compare to WT in Fig 2-7A). Mael−/− testis is smaller and far less complex than WT, and does not contain cells beyond early pachynema, the time of spermatogenesis arrest in this mutant. Very small cells (like elongated spermatids) and very large cells (like D spermatocytes and MI cells) are therefore absent from Mael−/− testis and do not contribute to the overall flow cytometry profile. On the Ho342-blue plot (**Fig 2-9B**), the 1C peak corresponding to haploid cells is completely absent (compare to WT in Fig 2-7C). Also observe the lack of the bimodal 4C peak, as compared to wild-type profile in Fig 2-7C. The bimodal distribution of cells with 4C DNA content is characteristic of the presence of L/Z and P/D populations. Disappearance of the bimodal peak is consistent with the presence of L/Z and absence of P/D populations.*

3. Draw gates around the major populations on Ho342-blue/Ho342-red plot as de- scribed (see Basic Protocol 3, step 2).

*Upon cell sorting, Mael−/− gate R6 contains Spg cells, gate R7 contains PL cells, and gate R2+R3 contains L and Z cells (**Fig 2-9C**), all like WT. However, unlike WT, the population within the R2+R3 gate also contains atypical cell types with a mix of LZ or ZP characteristics. These aberrant cells are characteristic of the reported mutant phenotype [116].*

4. View gates the originally drawn on the Ho342 plot (**Fig 2-9C**) on the FSC/SSC plot (**Fig 2-9D**), and then draw back-gates on the FSC/SSC plot to minimize contamination from adjacent or other populations, where appropriate (**Fig 2-9D**, R6 gate with Spg back-gate displayed only).

Note the separation of the R6 gate into two populations, one expanding along the FSC axis and the other expanding along the SSC axis. This separation suggests the presence of two different populations. Indeed, the extension of the R6 contour profile to the right (towards increasing SSC) indicates the presence of somatic cells. However, upon sorting, if the Spg back-gate has been set accurately, there should be no contamination of the Spg fraction with soma, since these appear distinct based on the scatter profile.

5. On the FSC/SSC plot, examine the other gates originally drawn on the Ho342 plot (**Fig 2-9C**) to establish whether back-gates need to be drawn.

Note the polarity of the resulting R2+R3 density, indicated by the two-lobed contour in Fig 2-9E. As in WT, the polarity of this contour is consistent with the presence of two populations, with the left lobe containing more L than Z spermatocytes, and the right containing more Z spermatocytes. In addition, the right lobe contains more of the aberrant Z cells and unusual cells having both Z and P characteristics.

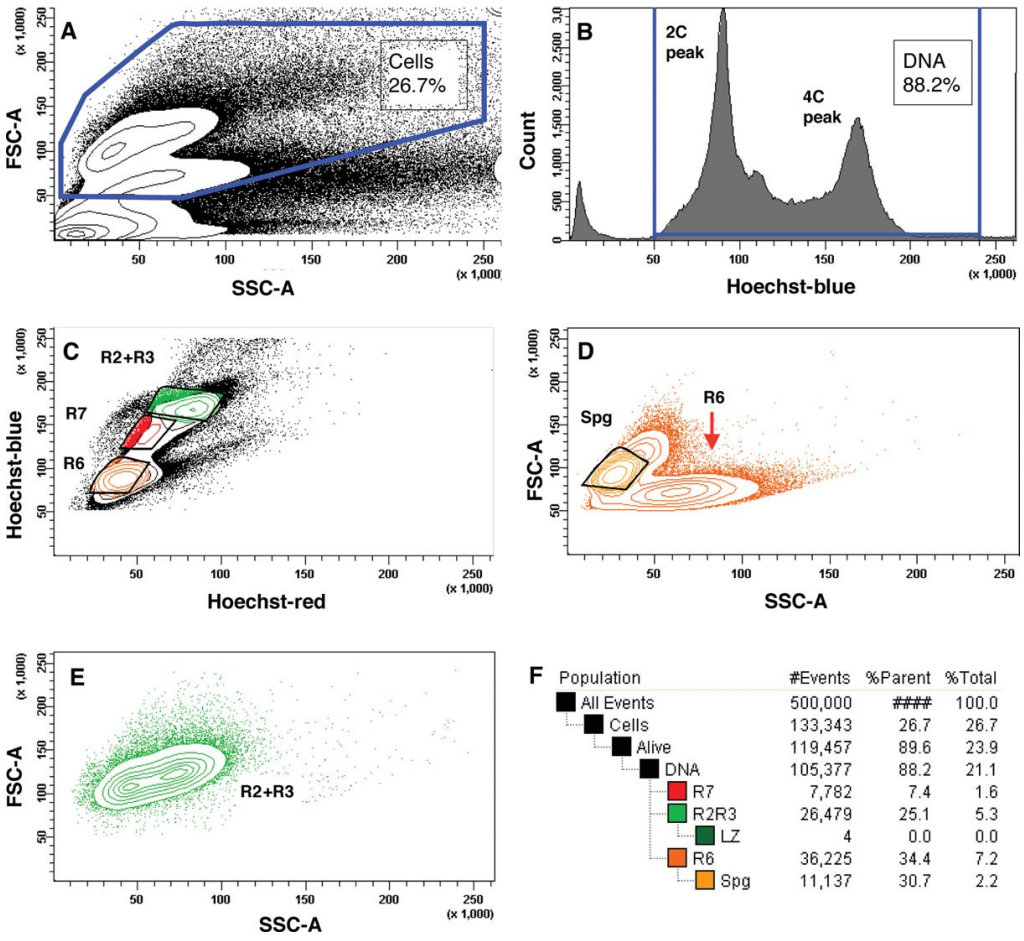


Fig 2-9 Gating and back-gating strategies for isolating individual MPI populations from a *Mael−/−* testicular cell suspension. All signals were acquired with linear amplification. Signals are plotted as area integrated intensity values. Percent values on plots represent percent of parent population (the latter identified below). Gate name is specified above the percent value. **A**) Debris exclusion is based on low light scattering parameters. Cells (inside blue gate) are distinguished from debris (outside blue gate) based on low light scatter characteristics viewed on the FSC/SSC contour (probability) plot. The parent of the "Cells" gate includes all cells acquired (500,000 cells). Observe that the light scatter profile of *Mael−/−* cells exhibits very few cells (outside major contour lines) with relatively high FSC characteristics (>150,000 units). This is apparent when comparing to WT (Fig 2-6A). **B**) DNA content based on Hoechst-blue fluorescence. Note

that the *Mael−/−* Hoechst-blue profile lacks the 1C peak and the bimodal 4C peak observed in WT mice (compare to C). This is because Mael−/− mice arrest before pachynema is complete, and thus lack the second part of the 4C peak that is characteristic of the P and D spermatocyte population. Populations that fall within the blue DNA gate include cells with 2C through 4C DNA content. Low Hoechst-blue events are eliminated and correspond to debris and dead cells. The parent of the "DNA" gate is the "Alive" gate from Fig 2-6B. **C**) Gating on individual testicular populations of interest based on Hoechst fluorescence. Colored regions represent gates drawn based on Hoechst-blue/Hoechst-red fluorescence. Gates shown: R6, R7, and R2+R3 containing spermatogonia (Spg), pre-leptotene (PL) spermatocytes, and a mix of L, Z, and aberrant cells, respectively. **D**) Back-gating approach to enrich for Spg. "Spg" is a back-gate of the R6 gate drawn on Hoechst fluorescence in panel A. The two shifts of the R6 population, one towards increasing FSC and the other towards increasing SSC, suggest the existence of two populations. Indeed, the population with the SSC shift indicates contamination with somatic cells present in this area of the FSC/SSC plot, and the "Spg" back-gate eliminates the contribution of this contaminating population from analysis and sorting. **E**) View of R2+R3 gate (originally drawn on the Hoechst fluorescence plot; panel C) on the FSC/SSC plot. Observe the polarization of the R2+R3 gate, and the formation of two lobes, indicative of the presence of two populations. **F**) Final gating tree indicating the sequential gating and back-gating procedure applied to the Hoechst/PI-stained testicular single-cell suspension before cell sorting.

SUPPORT PROTOCOL 1: Preparation of stained single-cell suspensions from adult murine testis

This protocol describes testis isolation, seminiferous tubule digestion into a single-cell suspension, and Ho342 and PI staining for FACS. Because of the histological organization of the testis, as well as the anatomical organization of germ cells within the testis proper, both enzymatic and mechanical dissociation methods are necessary to achieve a single- cell suspension that contains good amounts of the different types of testicular germ cells. Additionally, enzymatic digestion is critical for removing testicular somatic cells from the germ cells and testicular tissue, thus minimizing the somatic contribution to the overall testicular cell profile. This protocol has been described previously [28] and is presented here with minor adjustments. For additional or alternate information on preparation of adult murine testicular single-cell suspensions and Ho342 staining, refer to the following references [28, 78-83].

These steps describe the preparation of suspensions from both WT and *Mael−/−* knock-out mice. Spermatogenesis-defective testes often share fundamental characteristics, in- cluding small testicular and a significant reduction in cell number and complexity. Thus, the sample preparation described here should be applicable to other types of spermatogenesis-defective testes.

Materials

Wild-type male mice (C57BL/6 J), 2-4 months old (Jackson Laboratory)
Mutant male mice (*Maelstrom* knockout, *Mael−/−*), 2-5 months old [116]

Gey's balanced salt solution (GBSS; Sigma)
Collagenase/DNase solution (see recipe), ice cold
Trypsin/collagenase/DNase solution (see recipe), prewarmed (35°C)
2.5% (w/v) trypsin (10×) (Gibco)
1 mg/ml DNase I (Invitrogen, cat. no. 18068-015) in 50% (v/v) glycerol
10 mg/ml Hoechst 33342 (Ho342, aqueous solution; Life Technologies)
0.4% (w/v) trypan blue Newborn calf serum (NCS)
1 mg/ml propidium iodide (PI; Sigma)

Dissection tools: scissors, straight and angled fine-tipped forceps
15-ml sterile conical tubes (BD Falcon)

35°C shaking water bath or incubator
Plastic disposable transfer pipets
100- and 40-μm nylon cell strainers (Falcon)
Hemocytometer
Phase-contrast microscope
Sterile 5-ml polypropylene culture tube
12 × 75−mm tube with 35-μm cell strainer cap (BD Falcon)

Isolate testis

1. Sacrifice one WT or two *Mael−/−* adult male mice according to institutional animal care guidelines.

Use one WT mouse each for an unstained control, a PI-stained control, a Ho342-stained control, and a Ho342/PI-stained control. Use two mutant mice for each experimental sample, because the Mael−/− testis is smaller and has fewer cells. Prepare and run each control or experimental sample separately (not in parallel) so that cells can be sorted immediately after staining.

2. Remove WT testis (or both *Mael−/−* testes) and place in a Petri dish on ice con- taining 1 ml ice-cold GBSS.

3. Use a pair of forceps to remove the tunica albunigea.

4. Transfer each decapsulated WT testicle into a separate 15-ml tube on ice containing 6 ml collagenase/DNase solution. For *Mael−/−* mice, transfer all four testicles into a single tube containing 6 ml of the same solution.

Digest testis

5. Shake tubes in a horizontal position at 120 rpm for 10 min at 35°C. Halfway into the incubation, gently pipet the testicles up and down two or more times to facilitate tubule dispersion.

The temperature and agitation speed are the same for all incubation steps, unless noted otherwise. All pipetting steps should be performed using disposable plastic transfer pipets. At the end of incubation, tubules should be thin and dispersed, but not extensively sheared.

6. For WT samples only, stand the tube vertically and allow tubules to settle for 2 min at room temperature. Remove the supernatant (enriched in interstitial testicular cells), leaving just enough liquid to cover the settled tubules.

Although many interstitial somatic cells are removed with the supernatant, a large population still remains and will appear in the final digest. This step is omitted with the smaller mutant

testes because of the risk of losing germ cells along with the somatic cells.

7. Add 6 ml prewarmed (35°C) trypsin/collagenase/DNase solution to the tube and gently pipet the tubules up and down ten times to aid tubule fragmentation. Incubate 12.5 min at 35°C.

8. Add 60 µl of 2.5% trypsin, pipet ten more times, and continue incubating for another 12.5 min.

The tubules should appear fragmented and the solution dense with cells. If necessary, additional pipetting and gentle periodic tube inversion can aid tubule fragmentation.

9. Pipet the final suspension ten more times and pass through a 100-µm nylon cell strainer.

Prestain cells with Ho342

10. Set aside a 100-µl aliquot of cells for counting, then proceed immediately with Ho342 prestaining.

11. Add 10 µl of 1 mg/ml DNase I to the remaining suspension, followed by 10 µl (for WT) or 5 µl (for *Mael−/−*) of 10 mg/ml Ho342. Pipet up and down ten times and then incubate for 20 min, pipetting again halfway into the incubation.

For Mael−/− cells, it may be necessary to increase (e.g., double) the amount of DNase I here and in step 14 in order to eliminate clumps, because mutant testis contains many dead cells.

12. While cells are prestaining, dilute the 100-µl aliquot of cells (step 10) as follows, then count the number of cells in a 10-µl aliquot on hemocytometer under a light microscope. Count live cells based on trypan blue exclusion. For WT: dilute 1:6 by adding 200 µl GBSS and 300 µl of 0.4% trypan blue For *Mael−/−*: dilute 1:2 by adding 100 µl of 0.4% trypan blue.

Alternatively, WT cells may be diluted 1:4 by adding 100 µl GBSS and 200 µl of 0.4% trypan blue. Assessing accurate cell numbers is important for subsequent staining with Ho342 and for achieving a proper dye-to-DNA ratio, critical for good resolution of Ho342-stained testicular cells on a Ho342-blue/Ho342-red plot. See discussions on Ho342 fluorescence and DNA-binding [117] and the importance of Ho342 concentration [101]. One affect of too much Ho342 is excessive cell death and formation of clumps/debris in the suspension. Loss of cells and over-staining will change the final overall Ho342-stained cellular profile and decrease the overall resolution. On the other hand, too little Ho342 results in under-staining and poor separation of individual populations. The importance of the ratio of DNA-binding fluorochrome concentration to the number of binding sites (DNA content, cell number) in equilibrium staining obeys the law of mass action and is described UNIT 7.2 [118].

13. When prestaining incubation is complete, add 600 µl NCS to inactivate the trypsin and pipet

up and down five times.

Stain cells with Ho342

14. Add 10 µl of 1 mg/ml DNase I followed by 10 mg/ml Ho342 at a final concentration of 6 µg Ho342 per million cells. Pipet up and down ten times and incubate 25 min, pipetting again at the halfway point and the end of incubation.

15. Filter suspension through a 40-µm nylon cell strainer and keep on ice until sorting and analysis (up to 30 to 60 min).

Stain cells with PI and perform sorting

16. Prior to sorting, add 5 µl of 1 mg/ml DNase I to the Ho342-stained suspension and pipet up and down gently.

Addition of DNase I immediately before sorting is important for disaggregating clumps, which may otherwise clog the flow cytometer nozzle.

17. Transfer 2 ml suspension to a sterile 5-ml polypropylene culture tube and add PI stock solution to a final concentration of 2 µg/ml. Incubate 5 min at room temperature.

Typically, no sample dilution is necessary before sorting. The final concentrated samples (50 to 100 million cells, depending on age of the mouse) are usually at a good concentration for processing at a flow rate of 2000 to 3000 events/sec. A lower flow rate may be achieved by diluting the sample to 1−2 × 10^6 cells/ml with GBSS.

18. Filter cells through a 35-µm cell strainer cap into a 12 × 75−mm tube and keep on ice until analysis. FACS should be initiated within 30 to 60 min of this step.

SUPPORT PROTOCOL 2: Analysis of purity of sorted cells by immunofluorescence on nuclear spreads

After isolation of cells by FACS, they can be verified by monitoring the expression of meiotic and premeiotic germline markers. This protocol describes a procedure for labeling flow-

sorted cells with antibodies against SYCP3, γH2AX, and DMRT1. SYCP3 is a structural component of meiosis-specific synaptonemal complex (SC). Since SC develops gradually, its accumulation and differential distribution uniquely characterize a particular stage of MPI. γH2AX is a phosphorylated form of histone H2AX, which also shows differential and dynamic staining throughout MPI. Typically, meiotic germ cells can be distinguished from premeiotic Spg by the presence of SYCP3 and by unique γH2AX staining patterns. However, due to the presence of many different types of Spg, and due to the similarity between differentiated Spg and pre-leptotene cells, staining of sorted Spg and PL fractions for DMRT1 is also recommended. DMRT1 is a transcription factor that is present in Spg (both undifferentiated and most differentiated), but absent starting from the PL stage of spermatogenesis and onwards [119, 120]. The protocol below is performed according to Gaysinskaya et al. (2014) with minor modifications. Nuclear spreads are prepared as described by Peters et al. (1997) with minor modifications.

Materials

See Chapter 2.1 for details. Additional materials used here include mouse anti-DMRT1 primary antibody (200 μg/ml Santa Cruz Biotechnology, cat. no. H0712).

1. Analyze staining on a fluorescence microscope to score sorted fraction purity.

Typically, at least 50 cells are used to quantify the purity of Spg, Z, P, and D germ cells, and at least 75 cells are used to quantify the purity of PL and L germ cells. Extensive characterization and classification of sorted cells based on the intensity and distribution of DAPI, yH2AX, and SYCP3 signals are described elsewhere [27, 28, 104-106]. Analysis of MPI cells from sorted PL, L, Z, P, and D fractions is presented in Figs 3 & 4 [28]). Quantification of percent purity based on immunofluorescence after sorting can be found in [28]. Analysis of premeiotic Spg cells is presented in Fig 2-10.

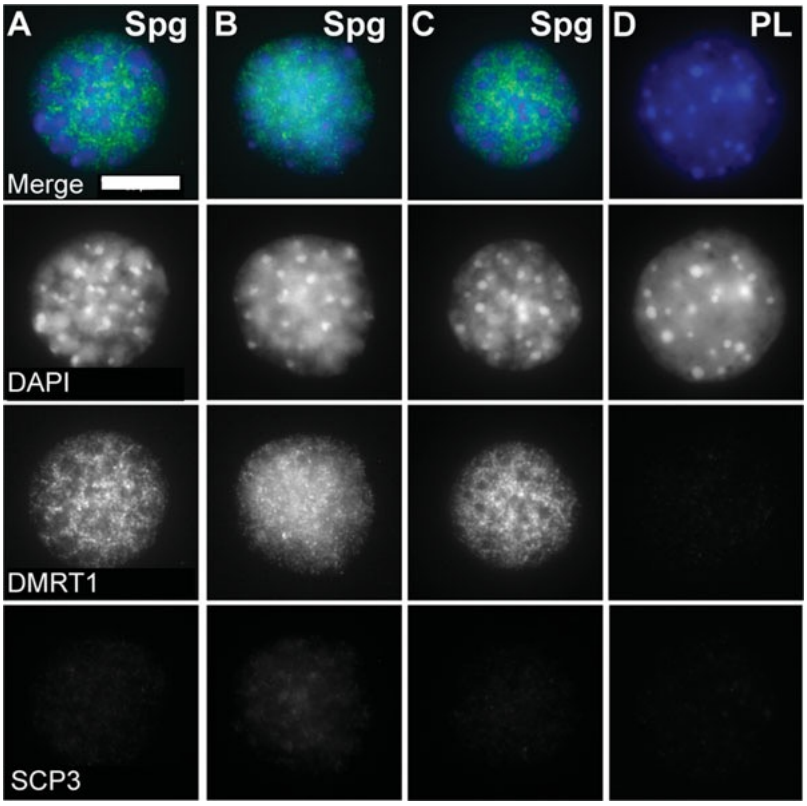


Fig 2-10. Immunofluorescence characterization of spermatogonia (Spg) after FACS. (A-C) Spread nuclei show examples of sorted Spg positive for DMRT1, a gonad-specific tran- scription factor expressed in all undifferentiated and most differentiated Spg. Different expression patterns and intensities of DMRT1 are consistent with DMRT1 expression in different types of Spg. **(D)** Pre-leptotene (PL) spermatocytes are negative for DMRT1. In all panels, the DMRT1 signal was enhanced for illustrative purposes using the Fiji package. Scale bar, 10 μm.

COMMENTARY

Background Information

Ho342-based dye staining protocols for enrichment of murine MPI populations have been previously reported [28, 79, 81, 82]. These and similar protocols vary in the degree of enrichment for individual MPI populations. More often than not, populations directly adjacent to one another on a Ho342-blue/Ho342-red plot are purified as a mix of leptotene/zygotene (L/Z) and/or pachytene/diplotene (P/D) spermatocytes [81-83, 121]. In addition, other testicular cell populations (including Spg and post-MPI cell types, such as haploid spermatids or meiosis II cells) show varying degrees of resolution on Ho342 plots [79, 81-83, 121]. Moreover, differences in testicular Ho342 fluorescence profiles observed between laboratories are likely due not only to differences in instrumentation, configuration, and settings, but also to differences in sample preparation, especially Ho342 staining methods. Nevertheless, Ho3342-based discrimination of individual MPI populations is a dominant and successful method for enrichment of highly pure testicular cell populations of interest, and is a critical tool for discriminating MPI cell types. The latter is especially important due to the absence of markers that can uniquely and unambiguously differentiate MPI subpopulations. For instance, although the synaptonemal complex protein SYCP3 exhibits differential pattern and intensity of staining, it is present in L, Z, P, D, and up to MII spermatocytes.

Critical Parameters and Troubleshooting

A number of references address critical aspects and difficulties in analysis and interpretation of flow cytometry data. Some of them focus on resolving particular problems, and others on general aspects of flow cytometry. Examples that fall into the first category include advice and guidelines for flow cytometry analysis of DNA content (e.g., [118]) or side populations [122]. Examples within the second category may, for example, provide guidance to improve data representation and visualization [115]. Several critical aspects of this protocol related to cytometer settings and

flow cytometry data display are addressed below.

Cytometer settings

To set up a flow cytometry experiment with specific fluorophores, the user must know what lasers are available for excitation, and what mirrors and filters are available for detection. Collectively, this is part of the optics configuration, which forms a system of light illumination and detection. The optics configuration must be set up with particular fluorophores in mind. For example, to excite Ho342, a cytometer equipped with a 375-nm near-UV laser or UV or violet lasers can be used. For detection, bandpass filters are used to filter light that is close to the emission peak of the fluorescent dye used (e.g., 461 nm for Ho342). For a 450/40 filter, the spectral band transmit- ted to the detector is 430 to 470 nm (450 ± 20 nm). Thus, one way to detect Ho342-blue is to use a 375-nm laser and a 450/40 bandpass filter in front of the PMT. In addition to band- pass filters, longpass filters are used to transmit wavelengths of light equal to or longer than a specified wavelength, and mirrors can serve as beam splitters to direct light of different wave- lengths in different directions.

An awareness of particular instrument specifications and settings can enable the use of a different but similar instrument, if necessary. In the authors' hands, similar profiles have been obtained from Ho342/PI-labeled testicular single-cell suspensions using two different instruments equipped with different Ho342-detecting lasers: a BD FACSAria III (BD Biosciences) equipped with a 375-nm laser, and a MoFlo cytometer (Dako Cytomation) equipped with a Coherent Enterprise II laser-emitting MLUV at 351 nm. Mirror and filter settings were adjusted in an instrument- specific manner. Using the BD FACSAria III, Ho342 detection was achieved by first splitting the Ho342 emission using a 610-nm dichroic longpass filter, and collecting Ho342-blue and Ho342-red emissions using a 450/40-nm bandpass filter and a 670-nm longpass filter, respectively. Using the MoFlo cytometer, Ho342 emission was also split using a dichroic mirror, and detection of Ho342-blue and Ho342-red was achieved using 450/65-nm and 670/30-nm

bandpass filters, respectively.

Data display

It is often best to choose contour plots over dot plots for data display. A contour plot (also called a probability plot) graphically displays relative frequencies of cells present at each point in the plot, thus providing a third dimension (frequency) to two-dimensional data. If probability is used to calculate contour levels, then the percentage of total number of events in outer and inner contour levels can be specified and adjusted for better visual view of populations of interest. In contrast to the dot plot, the contour plot becomes more accurate as the number of cells from which data are collected increases.

Biexponential plots are useful for accurate visualization of populations with low or background fluorescence, and are appropriate for viewing negative populations and drawing gates accurately around the major negative population. A great guide for FACS data pre- sentation is offered in Herzenberg et al. (2006).

Immunofluorescence analysis

DMRT1-staining of earlier (less differentiated) Spg can be much stronger than stain- ing in later (more differentiated) Spg. On meiotic nuclear spread preparations, DMRT1 immunofluorescence staining of differentiated Spg can be quite low and somewhat punctate. For example, DMRT1 staining of Spg that morphologically resemble PL spermatocytes can be very weak. Importantly, however, even weak DMRT1 staining seems to be highly specific, and it is unambiguously absent in control cells such as L, Z, P, and D cells, among others.

Anticipated Results

Based on immunofluorescence analysis after cell sorting, the purity of subpopulations falls within the following ranges: Spg, 80% to 91%; PL, 75% to 92%; L, 60% to 80%; Z, 75% to 90%; P,82%

to 95%; and D, 82% to 95% [28]. Often, the prevalence of L spermatocytes is lower than that of PL, Z, P, D, or Spg. Haploid cells can be enriched from the R0 fraction with a purity of >95% (**Fig 2-7A**).

Time Considerations

The fresh solutions needed to prepare testicular cell suspensions and to perform staining and immunofluorescence analysis can be made in ~30 min. Preparation of the cell suspension requires ~1.5 to 2 hr. Setting up the flow cytometer and cell sorter requires ~1 hr. Sorting gate adjustment, during which all gates and back-gates are checked for overlaps and adjusted, takes ~30 min. If necessary, PMTs are adjusted as well. Additionally, as mice of different ages may exhibit somewhat different prevalences of individual MPI cells relative to each other and to other cell types, the gates and back-gates may need adjustment to accommodate slight shifts in the Ho342 profile and variations in cell type abundance. Cell sorting can take up to 4 or 5 hr. The sample should be held at 4°C during sorting, and should typically stand on ice for no more than a total of 6 hr after preparation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material (doi: 10.1002/0471142956.cy0744s72).

Chapter 3

Transient Relaxation of DNA methylation in Meiotic Prophase I

Valeriya Gaysinskaya[1,2], Godfried W. van der Heijden[3], Brendan F. Miller[2,4], Kasper D. Hansen[5,6,7], Alex Bortvin[1*]

[1] Department of Embryology, Carnegie Institution for Science, Baltimore, Maryland, United States of America

[2] Department of Biology, Johns Hopkins University, Baltimore, Maryland, United States of America

[3] Department of Obstetrics and Gynaecology, Erasmus MC, University Medical Center, PO BOX 2040, 3000 CA, Rotterdam, The Netherlands.

[4] Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America

[5] Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, United States of America

[6] Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland, United States of America

[7] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, United States of America

Introduction

Cytosine DNA methylation is an epigenetic mechanism implicated in stable repression of Transposable Elements (TEs), X chromosome inactivation and monoallelic expression of imprinted genes [4, 123]. In mammalian cells, cytosine methylation occurs predominantly in the CpG context and is established by specialized enzymes known as DNA methyltransferases [56, 124]. Methyltransferase DNMT1 acts on hemimethylated CpG substrates and plays a key role in the maintenance of DNA methylation following the replication of the genome [55, 125]. In the absence of pre-existing DNA methylation, new cytosine methylation patterns are introduced by *de novo* methyltransferases DNMT3A and DNMT3B [55].

While DNA methylation is generally stable in differentiated somatic cells, preimplantation embryos and primordial germ cells (PGCs) experience genome-wide remodeling of DNA methylation [7, 46]. Global reduction of DNA methylation levels in PGCs is required for resetting of imprinting marks and acquiring totipotent developmental potential [52, 126]. Extensive remodeling of DNA methylation in PGCs also results in transcriptional changes, including expression of TEs [47, 74, 127, 128]. TEs are genomic DNA sequences that have the ability to become mobile and integrate into the genome [61]. TEs account for at least 40% of the mammalian DNA, are highly interspersed, and play a major role in shaping the mammalian genome structure, gene expression and epigenetic topology of chromatin [68]. TEs are classified by their mode of mobilization, either via a DNA (DNA transposons) or an RNA (retrotransposons) intermediate [61]. Although several families of retrotransposons co-exist in the genome, in placental mammals, including mice, Long INterspersed Element-1 (LINE-1 or L1) is the major category and contributes to over 20% of genome size [68]. While the vast majority of the estimated 600, 000 L1 elements in the mouse genome are old and mutated, and thus no longer capable of mobilization, a subset of intact, evolutionarily young copies could pose significant danger to the host by inserting into the genome, deleting and rearranging DNA or otherwise modifying the genome [62, 72]. To ensure robust transposon silencing, germ cells utilize a

specialized class of small RNAs, namely PIWI-interacting RNAs (piRNAs). In male fetal germ cells, piRNAs aid DNA methylation and repressive chromatin machineries to specifically repress transcription of young and potentially retrotranspositionally competent L1 loci. Transposon silencing is critical for normal differentiation of germ cells as abrogation of *de novo* DNA methylation and piRNAs in fetal male germ cells leads to high levels of L1 expression, accumulation of DNA damage and cell death in meiosis [129].

In a generally held view, once established in primordial male germ cells by the time of birth (through concerted efforts of *de novo* DNMTs and a their catalytically inactive accessory factor DNMT3L [126, 128]), DNA methylation levels remain stable and high during subsequent stages of spermatogenesis that take place after birth. However, this notion does not account for a long-standing observation of transient expression of L1 at the onset of meiosis as indicated by expression of L1-encoded protein ORF1p [94, 116, 130-133]. Although L1ORF1p expression in the wild type testis is weak when compared to that in various DNA methylation and piRNA mutant germ cells, the expression raised the possibility of transposon surveillance relaxation. A potential relaxation of transposon silencing at the onset of meiosis was hypothesized to be reflective of a change in DNA methylation [134].

Since L1 activity can be detrimental to the fitness of the host [62, 72], the coincidence of L1 expression in the WT during MPI is intriguing due to the potential of L1 to undermine key meiotic events of homologous chromosome synapsis and meiotic recombination [94, 116, 130, 133, 135]. Before entering MPI, preleptotene (PL) spermatocytes undergo a final round of DNA replication (meiotic S phase), which is much longer than the mitotic S-phase in the same organism [16-18, 20, 21]. MPI is also exceptionally long, lasting about 2 weeks [136], during which time the chromosomes undergo dramatic changes in their organization. MPI is subdivided into leptotene (L), zygotene (Z), pachytene (P), and diplotene (D) stages, based on their unique nuclear and chromatin morphology, although a change from one stage to the next involves rather gradual morphological transitions, but no proliferation [11]. After last DNA replication in PL

stage, the homologs begin to pair along their lengths and it is in the context of this unique pairing that key meiotic events occur. The initiation of homolog pairing takes place in L spermatocytes, with SPO11-mediated programmed meiotic double-strand breaks (DSBs) [137]. However, early homologous pairing may occur prior to DSB formation, during pre-meiotic S phase [27]. The L-to-Z (L/Z) transition is marked by a chromosome configuration called the "bouquet", where chromosome ends cluster and "integrate" into the nuclear envelope in a polarized fashion [86]. Homologous chromosomes become tightly paired at various points along their lengths in Z via prominent meiosis-specific structure, the synaptonemal complex. By P, the chromosomes synapse and complete meiotic recombination [138]. P is an extremely busy stage of MPI, exhibiting a dramatic increase in overall RNA and protein synthesis in preparation for the next phase of spermatogenesis [33]. At D, synaptonemal complex disassembles and homologous chromosomes de-synapse, remaining connected only at chiasmata, the physical remnants of homologous recombination [139].

To explore the molecular basis of L1 expression in wild type male meiosis we set out to precisely characterize the dynamics of L1 expression and to determine genome-wide DNA methylation levels of spermatocytes in the course of MPI. We find that DNA methylation levels of mouse spermatocytes are reduced at the onset of meiosis by replication-coupled passive DNA demethylation. In contrast to the accepted view of rapid DNA methylation maintenance within several minutes after DNA replication, we observe a gradual and slow process of DNA remethylation over the period of several days.

Results

Transient L1ORF1p expression in MPI of male germ cells

Evidence of L1ORF1p expression in WT meiosis is counterintuitive and puzzling since it is often a sign of concern [77, 116, 140, 141], however, it is also intriguing since it suggests a possibility of relaxed transposon control in wild-type MPI. Although several groups have seen L1ORF1p expression previously in male mice during MPI, most either attribute it to background levels in immunofluorescence staining or do not comment on the observed L1ORF1p expression at all [94, 116, 131-133]. One potential reason for lack of consistent reporting on L1ORF1p expression in meiosis is its examination relative to DNA methylation and piRNA mutants that exhibit excessive levels of L1ORF1p compared to the wild type. To provide a better temporal resolution of L1ORF1p expression during MPI in adult WT testis, we performed immunofluorescence analysis using the combination of anti-L1ORF1p antibody with well-characterized antibodies against acrosome-specific marker sp56 (**Fig 3-1)**, and DSB marker γH2AX for accurate staging of spermatogenesis (**Fig 3-2**) [142-145]. Spermatogenesis, which in mice takes ~ 35 days, is divided into three main parts, (1) the mitotic proliferation of Spg, (2) meiosis of spermatocytes and (3) spermiogenesis, or differentiation of spermatids into mature sperm. Spermatogenesis occurs within the seminiferous tubule of testis, where Spg are found at the basement membrane, but spermatocytes and spermatids are found progressively further from the membrane and towards the lumen, where mature sperm will be released (**Figs 1-2, 3-1 and 3-2**). The complex differentiation program involves many distinct steps resulting in an ability to distinguish the various types of undifferentiated Spg, differentiating Spg, meiotic and postmeiotic cells by their microscopical appearance [11]. In the testis of an adult mouse, cells at all stages of spermatogenesis are found simultaneously. Any given testicular cross-section will reproducibly account for a unique arrangement of up to four different cell types at different phases of spermatogenesis, with twelve possible permutations classified as the twelve stages (stages I through XII) of the seminiferous cycle (**Fig 1-2)** [11].

By evaluating L1ORF1p and other markers in the context of the seminiferous cycle, we established that L1ORF1p is reproducibly expressed during a particular window of MPI (**Figs 3-1 and 3-2**). Specifically, L1ORF1p expression begins approximately at the L/Z transition, persists through Z until mid-P but disappears in late P spermatocytes. L1ORF1p was also observed in round and elongating spermatids. In terms of the seminiferous cycle, the window of L1ORF1p expression in MPI spermatocytes corresponds to L/Z through mid-P spermatocytes found at stages X-VII (**Fig 3-1**). DSBs are a hallmark of Leptonema, and are progressively reduced (repaired) in Z, P and D. Using γH2AX staining in combination with L1ORF1p, we observe that L1ORF1p expression is indeed found in L/Z and Z spermatocytes, which are also positive for γH2AX (**Fig 3-2**). However, interestingly, in spermatocytes with most intense γH2AX staining, namely, in late L and L/Z transition [138, 145], no or extremely low L1ORF1p expression is observed, followed by more pronounced and intense expression in Z and early- and late- P spermatocytes. This result suggests that at the peak time of DNA nicks and breaks in early MPI, which is in theory a perfect opportunity for L1 insertion, WT meiosis seems to be protected from retrotransposition. Importantly, our analysis demonstrates that transient L1ORF1p expression is a characteristic feature of MPI in adult testis.

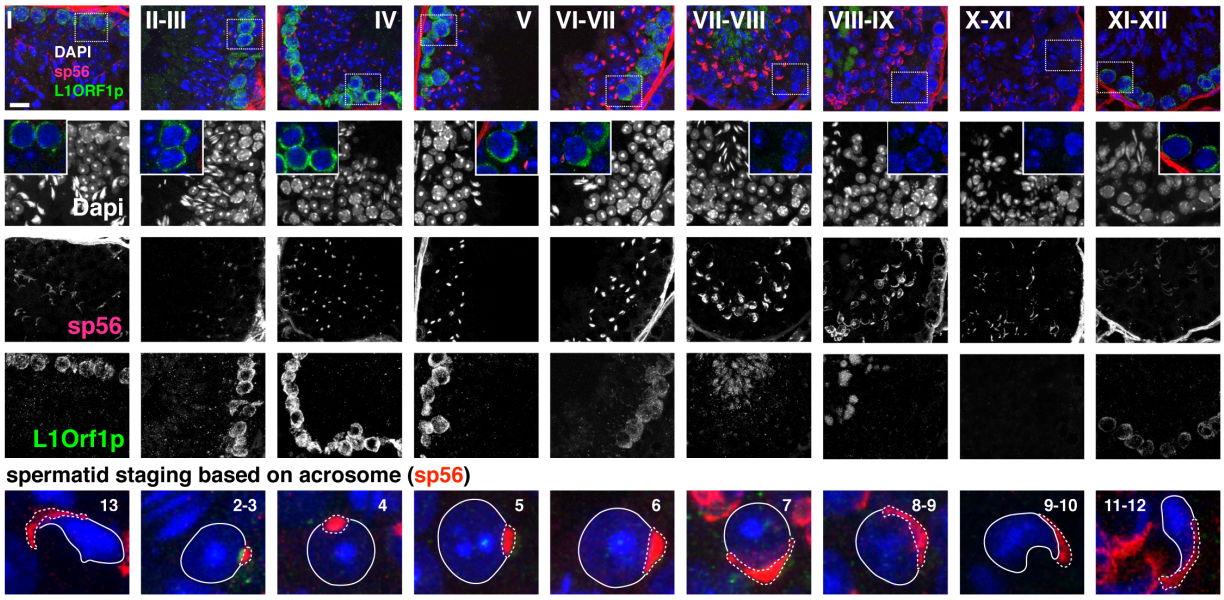


Fig 3-1. L1ORF1p expression in MPI of male germ cells. Temporal expression of L1ORF1p (green) was evaluated in testicular cryosections in the context of seminiferous epithelial cycle composed of stages I-XII. Haploid spermatids are identified based on numbers 1 through 16 according to degree of differentiation (only some are highlighted here). The basal membrane is outlined by the bright cross-reacting red staining. Following spermatogenic progression based on acrosome development marked by sp56 (red) and DNA stain, DAPI (blue), it is determined that cytoplasmic L1ORF1p is first seen in L/Z spermatocytes at stage XI (or X-XI) and persists from Z (stages XI-XII) to mid-P spermatocytes (stages I through VI-VII). L1ORF1p is not detectable in late P cells found stages VII-X, but is evident in the cytoplasm of elongating spermatids (see stages VII-IX) and is also detected as small dots in early round spermatids. The sp56 staining for spermatids beyond step 13 is difficult to see here, since the acrosome spreads very thin at this time. The selection inside the white box of the merged image (top row), is shown as a close-up inset in the DAPI-containing image row, and represents a single confocal plane in an otherwise 3-D stacked image, highlighting the cytoplasmic distribution of L1ORF1p in meiotic prophase I spermatocytes. Bar = 10 micron.

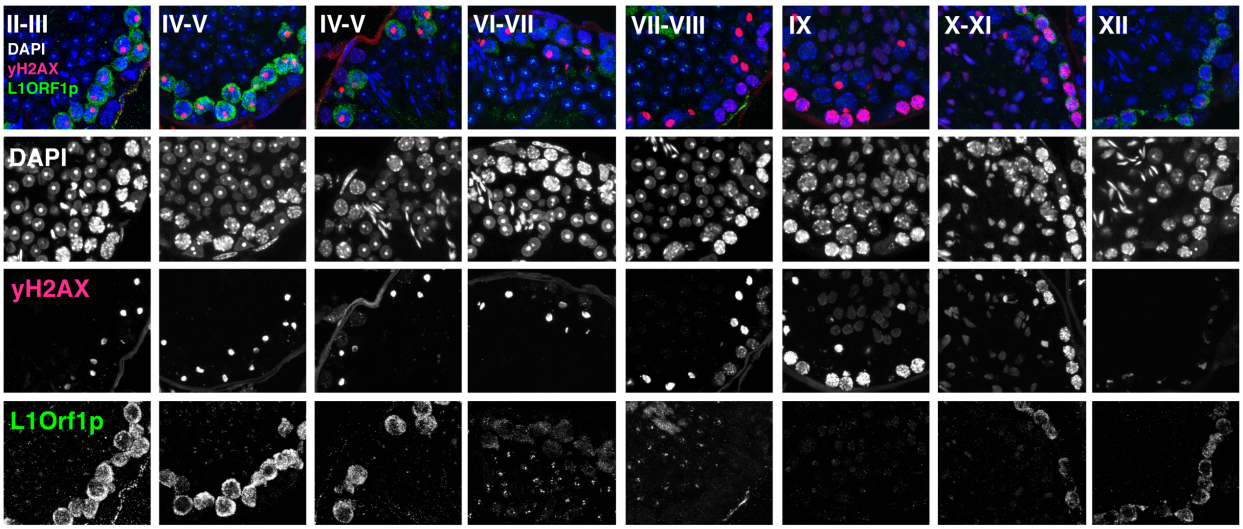


Fig 3-2. L1ORF1p expression in MPI in the context of meiotic Double Strand Breaks (DSBs). The composite image shows immunofluorescence staining of testicular cryosections with L1ORF1p (green) and phosphorylated histone variant and a DSB marker γH2AX (red), counterstained with DNA dye, DAPI (blue). The seminiferous epithelial cycle is marked by stages I-XII. During MPI progression, L1ORF1p is first seen in L/Z spermatocytes (seminiferous cycle stages X-XI), persists in Z (stages XI-XII) through mid-P spermatocytes (stages I through VI-VII). L1ORF1p is not detectable in spermatocytes at stages VII-X corresponding to tubules containing PL/L and late-P spermatocytes. L1ORF1p is also evident as small dots in the nucleus of early round spermatids, and in the cytoplasm of elongating spermatids. L1ORF1p is also detected as small dots in early round spermatids, especially evident in stages I-II.

Genome-wide DNA methylation analysis of adult MPI male germ cells

Given that L1ORF1p expression is indeed seen in wild type meiosis, we questioned the expectation of robust L1 silencing in MPI. Since DNA methylation is crucial for L1 repression, we posited if L1ORF1p expression in WT MPI is associated with hypomethylation of L1 elements. In our pilot experiments, we tested L1 DNA methylation by bisulfite conversion followed by PCR amplification with L1-specific primers and sequencing. Indeed, we observed decreased L1 DNA methylation in early MPI cells. This early observation led us to wonder if L1 demethylation is L1-specific or is a product of genome-wide demethylation. The latter idea relies on the fact that LINE elements are highly abundant and dispersed throughout the genome, and thus may serve as markers of other genomic events. To test these possibilities we investigated DNA methylation profiles, or methylomes, of individual MPI substages.

We enriched for individual MPI stages using Fluorescence Activated Cell Sorting (FACS) as previously described (**Methods and Procedures**) [28, 146]. We obtained two biological replicates for each stage: Spg, PL, L, Z, P, D and epididymal Spz germ cells from adult mouse testis. The purity of FACS-enriched MPI cells was verified by immunofluorescence with meiosis-specific (SYCP3, γH2AX) and Spg-enriched (DMRT1, DMRT6) markers as previously described [28, 146] (**Methods and Procedures**). For genome-wide analysis of DNA methylation at a single-CpG resolution, we performed Whole-Genome Illumina Bisulfite DNA Sequencing (WGBS) on genomic DNA extracted from FACS-enriched individual MPI stages. We used *Bismark* software for WGBS-Seq data alignment and methylation calling [147, 148]. On average, over 90% of the reads in each sample were aligned to the mouse reference genome (NCBI37/mm9) at least once (**Table 3-1, Protocol 1 in Chapter 4**). We spiked all our samples with an un-methylated phage lambda DNA to monitor bisulfite conversion efficiency, which resulted in more than 99.7% conversion rate for all samples (**Protocol 3-1 in Chapter 4**). Additionally, we used mouse epididymal sperm DNA (sequenced at lower coverage) as a highly methylated control. Each biological replicate accounted for 87-94% of all genomic CpGs in the

mouse genome, with average CpG coverage of 3x–6x per individual sample (**Table 3-2, Protocol 1**). Two biological replicates per data point exhibited high inter-individual Pearson correlation (**Table 3-3**) allowing for accurate comparison between the groups. Methylation levels at non-CpG sites, CHG and CHH were negligible (0.3-0.4%), thus we focused on CpG DNA methylation, which we will refer to hereafter as DNA methylation.

Sample ID	Sample Name	# Raw reads	# Failed to align reads	% overall read alignment	# Raw read pairs	# uniquely-aligned read pairs	% Uniquely aligned read pairs	# Non-uniquely aligned read pairs	# unique alignment pairs after de-duplication
VG1	WT Spg1	379,522,988	27,006,730	92.9%	189,761,494	145,868,488	76.90%	16,886,276	48,702,388 (33.39% of total)
VG3	WT PL1	375,506,566	25,605,205	93.2%	187,753,283	149,339,442	79.50%	12,808,636	43,558,443 (29.17% of total)
VG5	WT L1	330,089,994	30,845,183	90.7%	165,044,997	119,772,171	72.60%	14,427,643	75,420,707 (62.97% of total)
VG7	WT Z1	370,057,222	25,525,453	93.1%	185,028,611	141,091,120	76.30%	18,412,038	73,335,511 (51.98% of total)
VG911	WT P1	618,226,468	38,328,790	93.8%	309,113,234	243,331,265	78.7%	27,453,179	50,181,630 (20.62% of total)
VG12	WT D1	384,282,478	31,420,662	91.8%	192,141,239	141,691,298	73.70%	19,029,279	90,255,053 (63.70% of total)
VG2	WT Spg2	379,530,906	34,345,692	91.0%	189,765,453	138,826,170	73.2%	16,593,591	60930595 (43.89% of total)
VG4	WT PL2	356,090,204	32,699,710	90.8%	178,045,102	131,494,787	73.9%	13,850,605	69423704 (52.80% of total)
VG6	WT L2	367,267,278	34,300,532	90.7%	183,633,639	132,880,204	72.4%	16,452,903	45238124 (34.04% of total)
VG8	WT Z2	392,799,910	34,586,506	91.2%	196,399,955	143,329,093	73.0%	18,484,356	66470450 (46.38% of total)
VG10	WT P2	395,721,182	34,193,891	91.4%	197,860,591	144,366,629	73.0%	19,300,071	64755494 (44.85% of total)
VG13	WT D2	407,065,942	34,275,171	91.6%	203,532,971	149,180,976	73.3%	20,076,824	72023850 (48.28% of total)

Table 3-1. **Whole-Genome Bisulfite Illumina (WGBS) sequencing results.** Sequencing read alignment and de-duplication summary for wild-type (WT) samples. A read is said to align "uniquely" if it has exactly one valid alignment to the reference according to the alignment policy. Percent (%) overall alignment was calculated as (# of raw reads - # of failed to align reads)/total #

Sample ID	Sample Name	# M-bias filtered CpGs	Total # of covered CpGs	% total CpGs covered*	Average coverage of mappable CpGs	% Average CpG methylation
VG1	WT Spg1	70,741,691	19,255,878	88.6	3.7	81.0%
VG3	WT PL1	64,944,450	20,210,813	93.0	3.4	66.7%
VG5	WT L1	116,326,298	20,210,813	93.0	5.8	72.3%
VG7	WT Z1	110,225,074	20,137,088	92.7	5.5	76.8%
VG911	WT P1	69,520,427	18,968,431	87.3	3.7	82.2%
VG12	WT D1	144,980,758	20,491,322	94.3	7.1	81.3%
VG2	WT Spg2	92,941,328	19,967,351	91.9	4.7	78.8%
VG4	WT PL2	106,119,262	20,033,145	92.2	5.3	70.7%
VG6	WT L2	66,466,785	18,952,794	87.2	3.5	74.6%
VG8	WT Z2	97,986,908	19,947,972	91.8	4.9	76.2%
VG10	WT P2	93,713,906	19,696,249	90.7	4.8	82.1%
VG13	WT D2	113,948,241	20,271,739	93.3	5.6	81.7%
VG15	KO Spg1	110,969,098	20,199,547	93.0	5.5	74.8%
VG17	KO PL1	54,970,485	18,126,697	83.4	3.0	68.6%
VG18	KO PL2	102,164,107	20,064,801	92.4	5.1	70.3%
VG16	KO PL 3	100,052,651	20,040,071	92.3	5.0	67.1%
VG19	KO LZP-1	50,923,791	17,738,939	81.7	2.9	78.0%
VG21	KO LZP-2	83,302,813	19,641,331	90.4	4.2	76.1%
VG24	WT Sperm1	46,386,701	17,464,524	80.4	2.7	82.3%
VG25	WT Sperm2	52,583,228	18,011,652	82.9	2.9	82.8%

* Percent total CpGs covered is based on 21,722,957 CpGs in mm9 genome

Table 3-2. Summary of DNA (CpG) methylation evidence for deduplicated whole genome bisulfite sequencing reads that aligned uniquely to the mm9 mouse genome and overlapped at least once CpG cytosine. After read alignment and deduplication, Methylation bias (M-bias) across read length was evaluated and CpG methylation evidence was extracted using *Bismark*. Average CpG coverage and % CpG methylation were calculated using *bsseq* package (https://bioconductor.org/packages/release/bioc/html/bsseq.html).

A

Pearson correlation (r)	Spg-1	PL-1	L-1	Z-1	P-1	D-1	Spz-1
Spg-2	**0.982**	na	na	na	na	na	na
PL-2	na	**0.967**	na	na	na	na	na
L-2	na	na	**0.981**	na	na	na	na
Z-2	na	na	na	**0.988**	na	na	na
P-2	na	na	na	na	**0.984**	na	na
D-2	na	na	na	na	na	**0.995**	na
Spz-2	na	na	na	na	na	na	**0.981**

Table 3-3. Pearson Correlation between DNA methylation levels of biological replicates. CpG methylation for each sample was first averaged in bins of 500 CpGs (Methods and Procedures), and then Pearson correlation was calculated using *cor()* function in R.

Our analysis showed high median (> 84%) levels of CpG methylation in pre-meiotic Spg, P and D spermatocytes and haploid Spz (**Fig 3-3A, Table 3-4, Protocol 2 in Chapter 4**). This is consistent with a generally held view of high DNA methylation levels in adult spermatogenesis. Intriguingly, we uncovered an extended window of reduced global DNA methylation during early MPI demarcated by a pronounced drop in DNA methylation (~13pp) levels in PL followed by progressive gain of DNA methylation in L and Z spermatocytes, returning to pre-meiotic levels by P (**Fig 3-3A, Table 3-4, Protocol 2**). We estimate that the period of overall hypomethylation lasts up to 70 hours corresponding to the duration of PL, when hypomethylation occurs, and L and Z stages, when incomplete remethylation occurs. This analysis demonstrates the existence of transient relaxation of DNA methylation in PL, and suggests that DNA re-methylation occurs gradually, taking several developmental stages to reach Spg- or Spz-like levels.

Next, we wanted to know how DNA methylation is distributed on chromosomal and regional scales. To examine chromosome-wide distribution of DNA methylation in individual MPI stages, we summarized DNA methylation levels over a distance of 100 kb non-overlapping windows spanning the length of the chromosome. We found that global hypomethylation in PL is chromosome-wide (**Fig 3-3B**) and this was true for both biological replicates and all chromosomes examined. Thus, the drop in DNA methylation levels between Spg and PL is a genome-wide event that is not limited to certain chromosomes. This analysis also showed that the return of DNA methylation in MPI appeared predominantly as a chromosome-wide, gradual process.

To determine if DNA hypomethylation in PL is specific to a particular genomic annotation or is widespread, we examined the DNA methylation dynamics of exons, introns, intergenic and repetitive regions, as well as functionally specialized sequences such as promoters (+/-1000bp from TSS) and CpG Islands (CGIs) (**Fig 3-3C, Table 3-5**). Generally, DNA methylation levels at exons, promoters and CGIs (with average DNA methylation levels of < 70%, < 20% and <0.5% at all stages, respectively) are lower than average DNA methylation

levels at introns, intergenic regions and repetitive DNA. Similarly to the genome-wide DNA methylation profile, most of these genomic features are also subject to transient demethylation in PL (most prominently at introns, repeats and intergenic regions), except for CGIs whose methylation levels are already very low **(Fig 3-3C, Table 3-5**). The recovery of DNA methylation at these features also followed the general trend observed for the whole genome reaching pre-meiotic levels by P. The DNA methylation profile at promoters and exons exhibits an additional drop in DNA methylation levels between P and D, perhaps related to changes in transcription.

The global analysis of DNA methylation in MPI indicates that repetitive DNA exhibits comparable DNA methylation dynamics as non-repetitive genomic regions (**Fig 3-3C, Table 3-5**). Indeed, major classes of TEs, namely, the LINEs, SINEs, LTRs and DNA also experience transient demethylation in PL (**Fig 3-3D**). Our analysis suggests that L1 DNA becomes hypomethylated in PL preceding L1ORF1p expression in L or L-to-Z spermatocytes. This suggests that global genome-wide TRDM at meiotic onset may explain transient relaxation of transposon silencing (TRTS) in MPI [134].

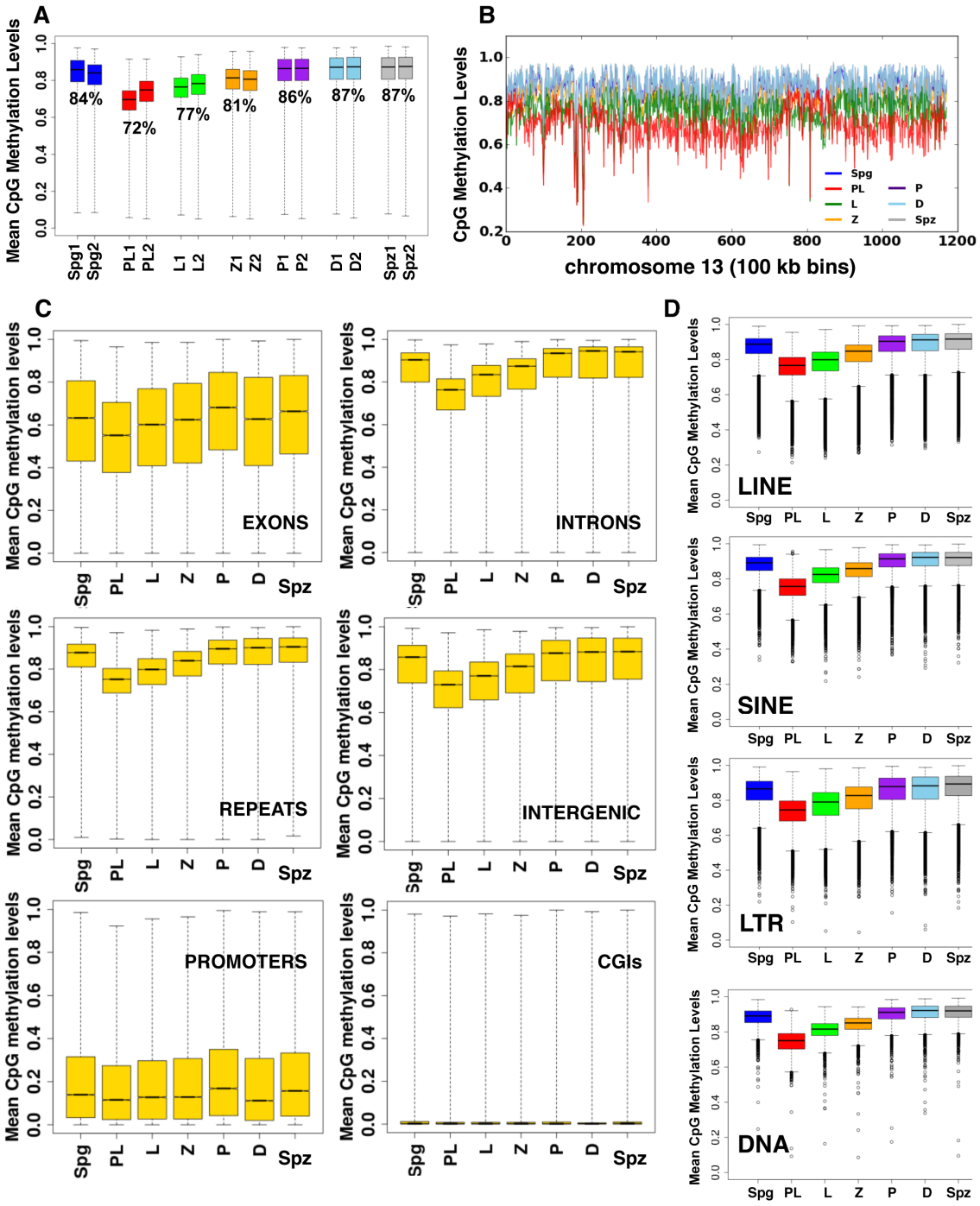

A
Mean CpG Methylation Levels
84%
72%
77%
81%
86%
87%
87%
Spg1 Spg2 PL1 PL2 L1 L2 Z1 Z2 P1 P2 D1 D2 Spz1 Spz2
B
CpG Methylation Levels
chromosome 13 (100 kb bins)
Spg
PL
L
Z
P
D
Spz
C
Mean CpG methylation levels
EXONS
INTRONS
REPEATS
INTERGENIC
PROMOTERS
CGIs
Spg PL L Z P D Spz
D
Mean CpG Methylation Levels
LINE
SINE
LTR
DNA
Spg PL L Z P D Spz

Fig 3-3. Global DNA methylation dynamics in MPI. (A) Global changes in DNA methylation reveal transient DNA demethylation in PL, followed by a gradual regaining of DNA methylation. DNA methylation was summarized as means of non-overlapping bins of 500 CpGs for individual biological replicates. Box-and-Whisker plot shows the maximum, upper quartile, median, lower quartile and minimum of data. Median percent DNA methylation for both replicates is specified underneath the boxplot. Spg (blue), PL (red), L (green), Z (orange), P (purple), D (light blue) and Spz (grey). **(B)** Chromosome-wide DNA methylation levels were plotted across chromosome length (chromosome 13, replicate 1 is shown). DNA methylation was averaged using sliding non-overlapping bins of 100 kbp. PL (red) exhibits the lowest overall DNA methylation levels, except at regions of "switching" in DNA methylation pattern, where L spermatocytes (green) exhibit the lowest DNA methylation levels instead. The levels in Spg, P, D and Spz are very similar, resulting in a blend marking the upper limit of the DNA methylation trace. DNA methylation of L and Z (orange) progressively increase. **(C)** Box-and-Whisker plot of DNA methylation levels across various genomic features. The average DNA methylation levels were aggregated as consecutive, non-overlapping averages of 100 CpGs. Averages were combined for biological replicates. **(D)** Box-and-Whisker plot of DNA methylation levels across different classes of transposable elements: LINE, Short Interspersed Nuclear Elements (SINE) and Long Terminal Repeats (LTR) retrotransposons and DNA transposons. The average DNA methylation levels were summarized as consecutive, non-overlapping averages of 100 CpGs and combined biological replicates.

Sample ID	Sample Name	Mean %CpG methylation, common (nonoverlapping 500 CpGs sliding windows)	Median %CpG methylation, common (nonoverlapping 500CpGs sliding windows)
AVG (VG1, VG2)	**WT Spg**	82.77%	84.83%
AVG (VG3, VG4)	**WT PL**	70.98%	72.19%
AVG (VG4, VG5)	**WT L**	76.06%	77.41%
AVG (VG6, VG7)	**WT Z**	79.21%	81.01%
AVG (VG6, VG7)	**WT P**	84.37%	86.53%
AVG (VG8, VG911)	**WT D**	85.11%	87.27%
AVG (VG24, VG25)	**WT Spz**	85.27%	87.44%

% of common CpGs covered across Biological Replicate 1 is 13882854 (64% of all CpGs in mm9 mouse genome)
% of common CpGs covered across Biological Replicate 2 is 15071422 (69% of all CpGs in mm9 mouse genome)

Table 3-4. **Genome-wide summary of mean DNA methylation.** CpG methylation was calculated for wild-type (WT) samples. CpG methylation was calculated as means of running non-overlapping bins of 500 CpGs. Mean value is reported here, and overall distribution is reported in Fig 2A.

Sample Name	EXON CpG methylation	INTRON CpG methylation	INTERGENIC CpG methylation	RMSK CpG methylation	PROMOTER CpG methylation	CGI CpG methylation
Spg	0.632	0.904	0.858	0.8778	0.1393	0.0033
PL	0.551	0.763	0.730	0.7525	0.1155	0.0025
L	0.601	0.835	0.771	0.7988	0.1275	0.0025
Z	0.624	0.874	0.815	0.8399	0.1283	0.0025
P	0.681	0.935	0.8766	0.8956	0.1684	0.0020
D	0.627	0.946	0.8823	0.9009	0.1118	0.0025
Spz	0.663	0.942	0.8837	0.90488	0.1571	0.0033

Table 3-5. **Summary of DNA methylation means across different genomic features.** DNA methylation corresponding to exons, introns, intergenic regions, repetitive regions annotated by RepeatMasker (RMSK), gene promoters and CpG Islands (CGIs) was calculated as means of running non-overlapping bins of 100 CpGs. Mean value is reported here, and overall value distribution is reported in Fig 3-2C.

We were motivated to examine DNA methylation dynamics of chromosome X (chrX), due to its unique meiotic behavior which includes transcriptional silencing in Z-to-P transition and heterochromatic condensation into an XY body, known as meiotic sex chromosome inactivation (MSCI) [30]. X chromosome inactivation in meiosis is regulated by DNA methylation and other epigenetic mechanisms [30]. We examined DNA methylation dynamics on chrX and compared it to two autosomes of comparable length, chromosomes 1, 2 and 3 and to all autosomes combined. The data show that chrX, like other chromosomes, exhibits transient DNA demethylation in meiotic onset followed by remethylation, although chrX tends to be hypomethylated at all stages (**Fig 3-4A**). Interestingly, DNA methylation levels in Spg-to-PL and PL-to-L transitions are distinctly less correlated to each other on chrX as compared to the autosomes, suggesting differences in the dynamics of chrX demethylation and remethylation (**Fig 3-4B**).

Cumulatively, using FACS-enrichment of MPI substages and genome-wide WGBS approach we were able to, for the first time, measure DNA methylation dynamics in individual stages of MPI, detect dynamic changes in wild-type DNA methylation levels in MPI and reveal existence of transient relaxation of DNA methylation (TRDM).

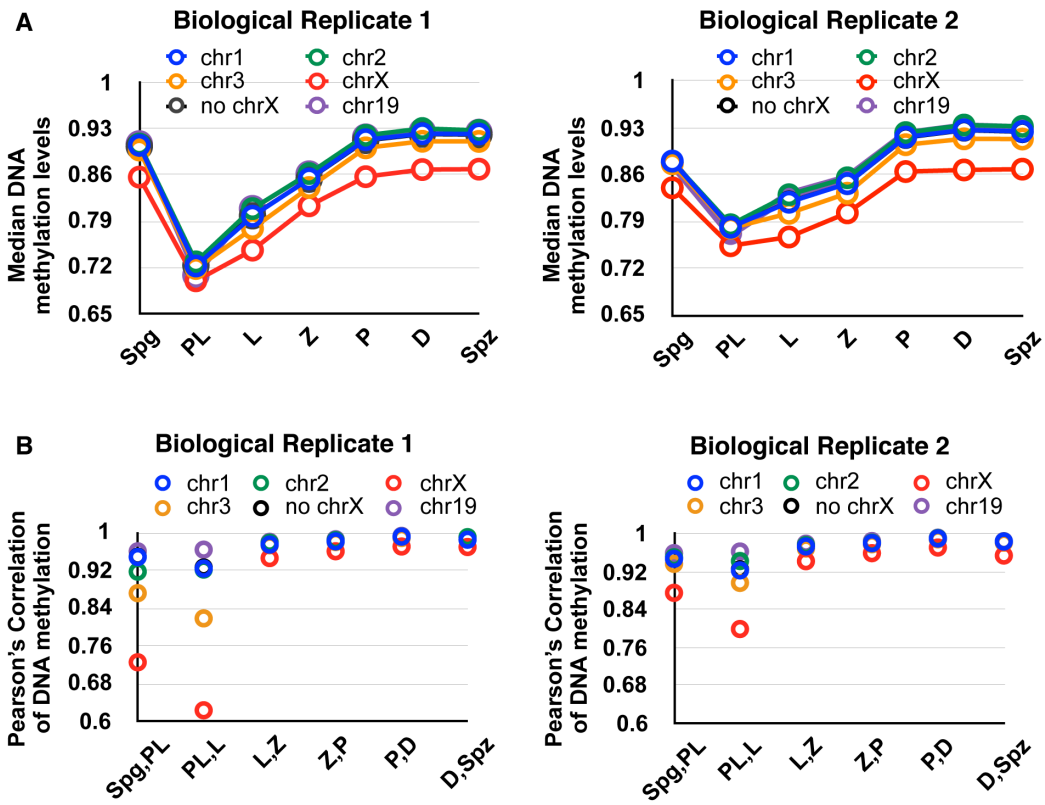


Fig 3-4. DNA methylation dynamics on chromosome X (chrX) compared to autosomes. Biological replicate 1 (left panel) and 2 (right panel): were examined independently. CpG methylation was averaged in bins of 100 CpGs **(A)**, and then Pearson correlation was calculated **(B)**. The number of different CpGs (CpG loci) evaluated for biological replicates 1 and 2 were as follows: chrX (214,981 and 266,439), chr1 (6,983,222 and 7,564,250), chr2 (1,022,879 and 1,103,479), chr3 (805,182 and 875,167), chr19 (383,441 and 412,451) all minus chrX (no chrX) (13,667,873 and 14,804,983).

Genome-wide DNA methylation pattern in PL overlaps with replication timing pattern

We wanted to understand the cause of transient hypomethylation in PL. A clue came from visual inspection of patterns of chromosome-wide DNA methylation levels (**Fig 3-3B and Fig 3-5A**). Focusing on PL DNA methylation pattern one can observe many regions of relative DNA hypomethylation interrupted by a few prominent regions of relative hypermethylation (**Fig 3-5A**, red trace, boxed region shows an example of a transition from hypomethylated to hypermethylated region in PL). We found that every chromosome, and both biological replicates possessed such prominent subchromosomal domains (**Fig 3-5A,** e.g. boxed region). The domains that show higher relative DNA methylation levels in PL, show lower DNA methylation levels in L, resulting in an apparent switch in DNA methylation traces in PL and L when compared to the rest of the chromosome (**Fig 3-5A**, e.g. boxed region and zoom, observe PL, red trace and L, green trace).

Some resemblance of this pattern to the Giemsa chromosome staining, or R/G banding pattern, led us to consider DNA methylation in the context of replication timing domains. Replicating domains are large-scale genomic territories that replicate at particular times during S-phase [149, 150]. Global early or late replication timing profiles appear relatively preserved between different cell lines and cell types tested, although there are many tissue-specific differences [150, 151]. Remarkably, an overlay of the chromosome-wide DNA methylation pattern from our data with replication timing domains (from a mouse CH12 cell line) (http://www.replicationdomain.com) revealed a strong overlap between the two (**Figs 3-5A and 3-5B**) [152]. Specifically, in PL cells, we observe an overlap between relative-hypermethylated DNA regions (**Fig 3-5A,** most prominent one is outlined) and late replicating domains (**Fig 3-5B**) (Figs **3-5A and 3-5B,** red trace). Correspondingly, an overlap is observed in PL between large-scale hypomethylated regions (**Fig 3-5A**, sample hypomethylated region is outlined) and early replicating domains (**Fig 3-5B**). Interestingly, a switch between DNA hypo- and hypermethylation in PL (**Fig 3-5A**, enlarged subset) is marked by an opposite switch in DNA

methylation pattern in L spermatocytes (**Fig 3-5A,** red and green traces for PL and L, respectively). This switch in DNA methylation pattern in PL-to-L transition is consistent with the genomic coordinates of the transition from early to late replication timing domains (**Fig 3-5B,** enlarged subset). The overlap between DNA methylation pattern and replication timing pattern in PL was true of both biological replicates examined independently and all chromosomes (see **Fig 3-6** for a another chromosome example).

Since our data suggests a link between DNA methylation in PL and replication timing we hypothesized that DNA hypomethylation in PL is related to replication. It has been shown that replication timing may be estimated from sequence coverage of unsynchronized cells [153]. Sequence coverage was also used to evaluate underreplication in Drosophila polytene chromosomes [154]. To this end, we evaluated the uniformity of genome sequencing coverage in our WGBS data. Specifically, we summarized read frequency over a distance of 5kb non-overlapping windows spanning the length of the chromosome and corrected for the difference in total read count between the samples (Methods and Procedures). Remarkably, in our analysis, we observe consistently lower sequencing coverage in the hypermethylated domains/late replication timing domains in PL (**Fig 3-5C**), disappearing in L, consistent with the lack of replication in L phase.

Together, these observations make us hypothesize that the observed hypomethylation represents hemi-methylated DNA after replication but prior to re-methylation. The observation that regions of the genome in early replication timing domains show hypomethylation earlier (in PL) than regions of the genome in late replication timing domains (L) is consistent with this hypothesis; note that the estimated length of mouse PL stage is ~2 days (including S-phase which is ~20 hours) and the length of L stage is 1-2 days [13, 136].

To further confirm PL cells we obtained by FACS-enrichment are replicative, we injected mice with EdU and after 2h performed FACS enrichment, followed by EdU detection of PL cell fraction, a population with a characteristic S-phase ploidy span of 2N-4N. Indeed, at least 70%

our FACS-enriched PL cells are EdU-positive, with the majority of EdU patterns corresponding primarily to middle and late S phase (**Fig 3-7**) [27, 155].

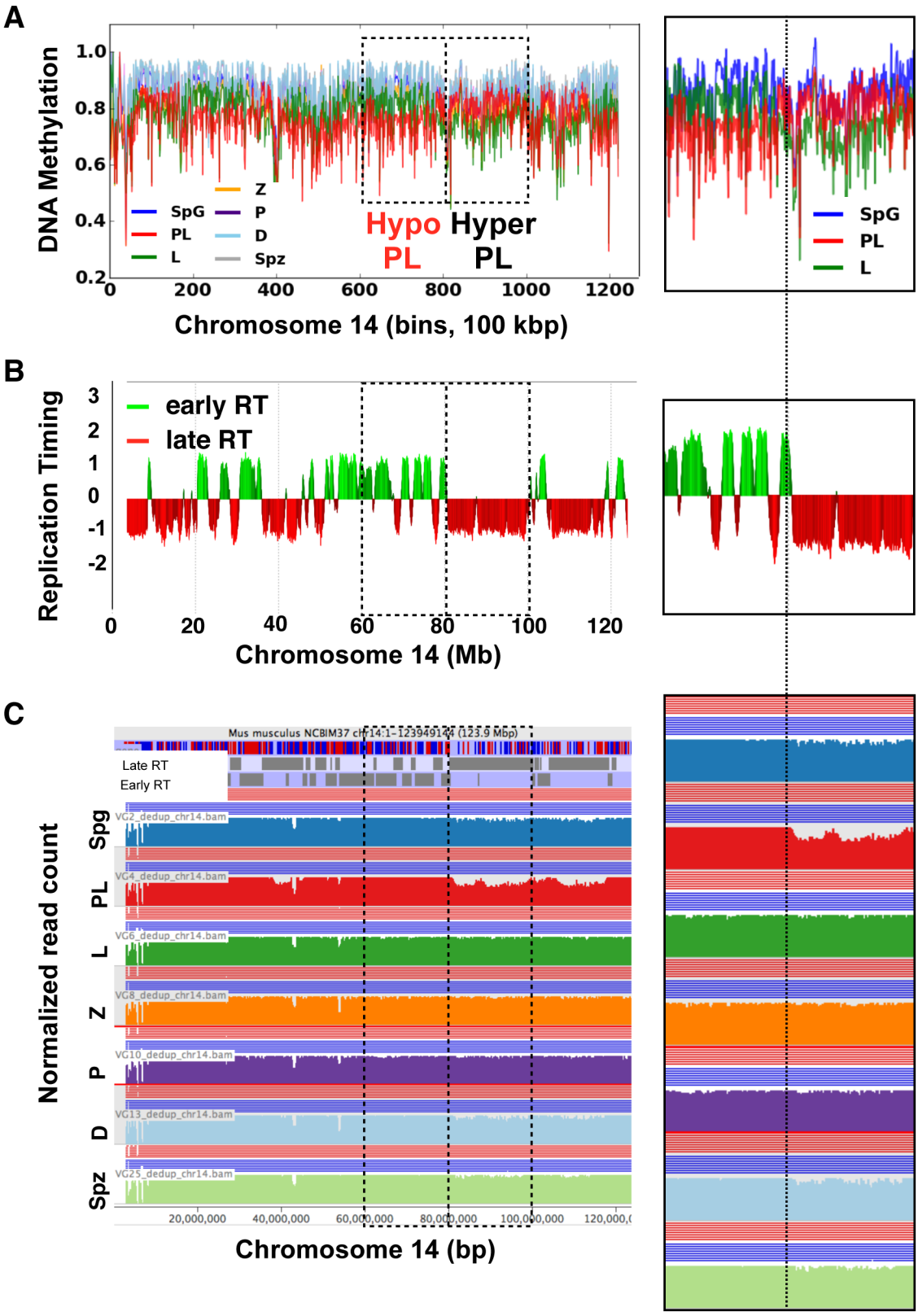

A
DNA Methylation
1.0
0.8
0.6
0.4
0.2
SpG
PL
L
Z
P
D
Spz
Hypo PL
Hyper PL
0
200
400
600
800
1000
1200
Chromosome 14 (bins, 100 kbp)
SpG
PL
L
B
Replication Timing
3
2
1
0
-1
-2
early RT
late RT
0
20
40
60
80
100
120
Chromosome 14 (Mb)
C
Normalized read count
Mus musculus NCBIM37 chr14:1-123949144 (123.9 Mbp)
Late RT
Early RT
Spg
PL
L
Z
P
D
Spz
VG2_dedup_chr14.bam
VG4_dedup_chr14.bam
VG6_dedup_chr14.bam
VG8_dedup_chr14.bam
VG10_dedup_chr14.bam
VG13_dedup_chr14.bam
VG25_dedup_chr14.bam
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,000,0
Chromosome 14 (bp)

Fig 3-5. DNA methylation pattern in PL overlaps with replication timing. (A) Plot of CpG DNA methylation of MPI stages, pre-meiotic Spg and post-meiotic Spz, across chromosome 14, which is ~125 Mbp long. Chromosome-wide DNA methylation was averaged using sliding non-overlapping windows of 100 kbp. The procedure was performed on every chromosome, and both replicates, independently. **(B)** Replication timing (RT) data from CH12 cells (mouse B-cell lymphoma) (http://www.replicationdomain.com) and **(C)** Genome sequencing coverage after WGBS, viewed in SeqMonk program. Shown, for each sample is the read count quantification of sequence coverage summarized as averages of sliding non-overlapping 5 kbp windows and corrected for largest data set.

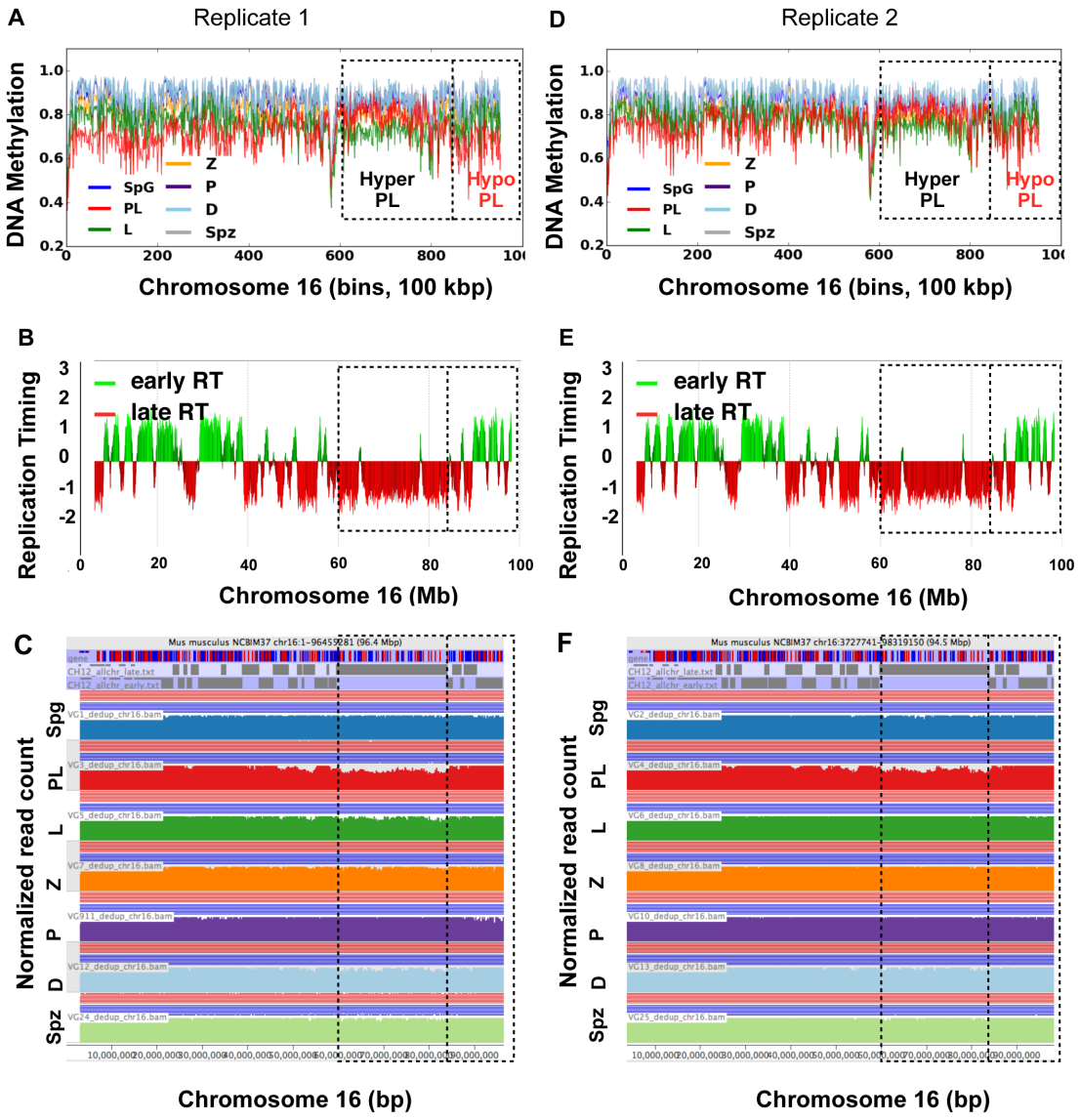


Fig 3-6. DNA methylation pattern in PL overlaps with replication timing, an example of chromosome 16. Biological replicate 1 is shown on the left, and replicate 2 on the right. (A,D) DNA methylation, (B,E) Replication Timing (RT) and (C, F) Genome sequencing coverage for two biological replicates. Replication timing (RT) data (in B and E) is from CH12 cells (mouse B-cell lymphoma) (http://www.replicationdomain.com). Genome sequencing coverage after WGBS-seq is viewed in SeqMonk program (in C and F).

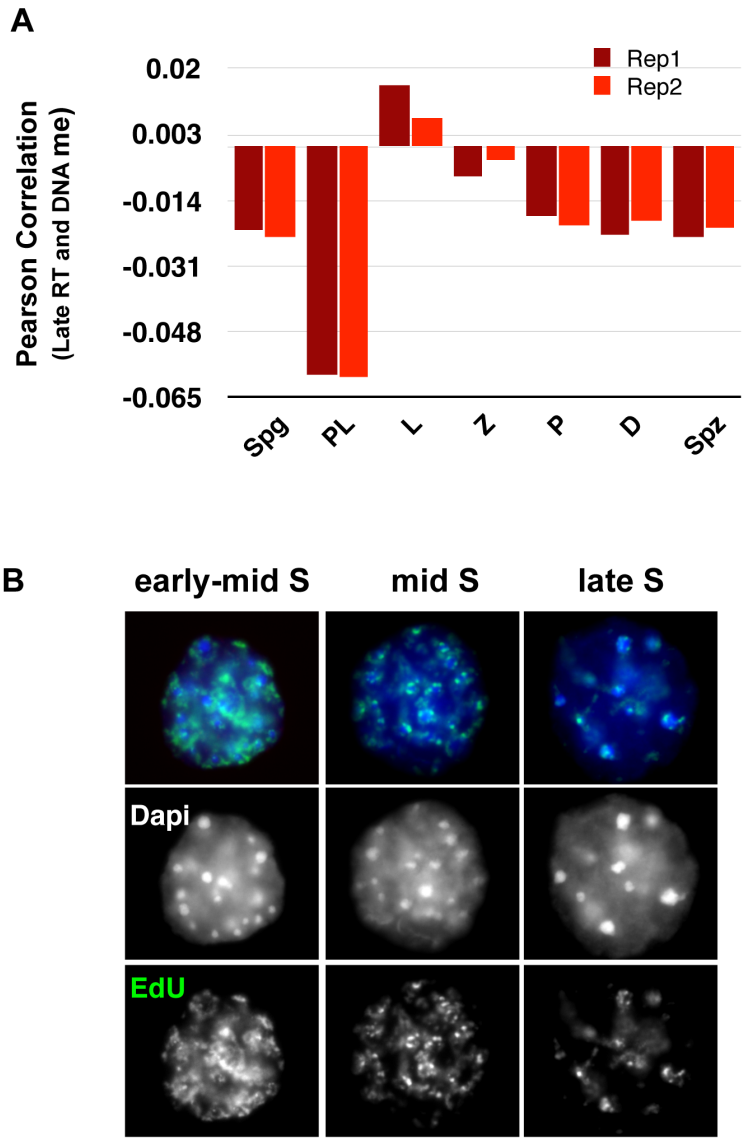


Fig 3-7. Relationship between replication timing and DNA methylation. (A) Replication timing (RT) data from CH12 cells (mouse B-cell lymphoma) (http://www.replicationdomain.com) was correlated with CpG DNA methylation corresponding to late RT domains. Note an existence of a prominent switch in correlation directionality from PL to L. Biological replicates (Reps) were processed individually and are shown in light and dark red. **(B)** The PL cell fraction enriched by FACS contains replicating cells. More than 70% of FACS-enriched PL cells are EdU+, enriched in mid- and late- S phase, based on the characteristic EdU staining patterns.

Differentially methylated regions are found throughout MPI

Given the global differences in DNA methylation between Spg and PL and throughout MPI, we wanted to better understand the remethylation dynamics and to examine if all of the changes in DNA methylation are due to TRDM in early MPI. Thus, we identified regions that exhibit significant differences in DNA methylation patterns between individual MPI stages and examined their dynamics during MPI. We used *bsseq* approach to identify differentially methylated regions (DMRs) between contiguous MPI stages in a statistically-principled, coverage-conscious and biological replicate-aware manner (**Protocol 3 in Chapter 4**) [156]. For DMR analysis we required that each CpG used for the estimation of DNA methylation be covered at least once in all 4 samples (2 biological replicates per 2 stages). This selection resulted in a median CpG coverage of 3X-7X per sample (or 6X-14X per biological duplicate) and an overall coverage of >77% of all genomic CpGs (**Table 3-6**).

WT Spg vs PL comparison	CpG coverage (median)	# CpGs covered (in common)	% Total CpGs evaluated (# CpGs in common/total # CpGs)
Spg1	3	16,825,193	**77.5**
Spg2	4		
PL1	3		
PL2	5		
WT PL vs L comparison			
PL1	3	16,774,643	**77.2**
PL2	5		
L1	5		
L2	3		
WT L vs Z comparison			
L1	5	17,641,230	**81.2**
L2	3		
Z1	5		
Z2	4		
WT Z vs P comparison			
Z1	5	17,343,570	**79.8**
Z2	5		
P1	3		
P2	4		
WT P vs D comparison			
P1	3	16,433,535	**75.7**
P2	4		
D1	7		
D2	5		

* Percent total CpGs is based on 21,722,957 CpGs in mm9 genome

Table 3-6. **Summary of CpG coverage for the analysis of Differentially Methylated regions (DMRs).** Only CpGs in common between all four samples of a pairwise comparison (two biological replicates per group) were included in the analysis.

We identified DMRs between Spg and PL, PL and L, L and Z, Z and P and P and D (**Fig 3-8A, Table 3-7**). More than 99% of all DMRs between Spg and PL are hypomethylated with respect to PL, and more than 98% of DMRs between Z and P become remethylated in P (**Fig 3-8A,B Table 3-7**). Relatively few DMRs are thus detected between P and D, cumulatively implicating only ~ 190 Mbp of DNA, as compared to ~ 1.3 Gbp of DNA implicated in DMRs formed between Spg and PL (**Table 3-7)**. Consistent with our earlier findings that in PL spermatocytes DNA hypomethylation is enriched in early replicating domains, we find that genome-wide, ~60% of all early replication timing domains overlapped with hypomethylated DMRs (**Methods and Procedures**). Conversely, in L, there is a pronounced enrichment of hypomethylated DNA in late replication timing sequences. Thus, DMR analysis is in good agreement with our replication timing analysis.

We used DMR analysis to elaborate on remethylation dynamics. Large hypomethylated DMRs present in PL become gradually remethylated in L and Z, thus, DMRs between these stages exhibit a progressive decrease in mean methylation difference (**Fig 3-8B, Table 3-7**) and become progressively smaller in size (**Fig 3-8C, Table 3-7)**. Indeed, the DMRs present between Spg and PL included up to ~56% of all CpGs evaluated, while those between PL and L, and L and Z included ~ 41% and ~3% of all CpGs, respectively (**Fig 3-8D, Table 3-7**).

Overall, these DMR dynamics primarily reflect demethylation in PL followed by remethylation of the same CpGs in L and Z. Indeed, by examining how many DMRs overlap with each other, we find that DMRs formed between Spg and PL account for up to 75% of all DMRs identified between PL and L and up to 63% between L and Z (**Methods and Procedures**). The gradual remethylation of hypomethylated DMRs in PL is also reflected by the decreasing genomic size of DMR blocks between PL and L, and L and Z as compared to the large DMR blocks between Spg and PL, a feature that results in no single DMR exactly matching the genomic coordinates of the original DMRs between Spg and PL (**Fig 3-8C, Table 3-7**).

Interestingly, although global levels of DNA methylation in Z are reflective of being remethylated relative to the preceding stages, Z is still hypomethylated relative to P (**Fig 3-2, Fig 3-8A,B**). Accordingly, during Z to P transition, we observe that up to ~ 57% of all CpGs in common between Z and P undergo a subtle burst of remethylation, from around 81% in Z to 88% P, suggesting that the bulk of remethylation that reaches pre-meiotic or almost Spz-like levels occurs between Z and P stages of MPI. Indeed, the original DMRs formed between Spg and PL, can explain up to 75% of all DMRs observed between Z and P (**Methods and Procedures**). We find that gradual remethylation concerns all genomic features examined (exons, introns coding sequences and repeats) (**Table 3-8B)** and, in Z, up to 60% of these features are still hypomethyalated compared to P, although mean DNA methylation difference at the hypomethylated DMRs is relatively small. In P less than two percent of these features are found in hypomethylated DMRs relative to D, due to remethylation.

Intriguingly, while P and D share very similar DNA methylation profiles overall (**Fig 3-3A**), a subset of DMRs is observed between them. These DMRs reveal hypomethylation from P to D that involves only up to 8% of all examined CpGs in common and relatively small mean genomic size (~12 Kbp, as compared to ~35 kbp in PL) (**Fig 3-8A-C, Table 3-7**). Considering that original DMRs formed between Spg and PL account for only 50% of all DMRs found between P and D, it is likely that the hypomethylation observed in late MPI (P-D) is unrelated to TRDM. Together, the above analysis suggests that genome-wide remethylation during TRDM is not uniform and results in existence of DMRs throughout MPI. The data also point to the emergence of TRDM-independent DMRs that could be associated with other events taking place during MPI.

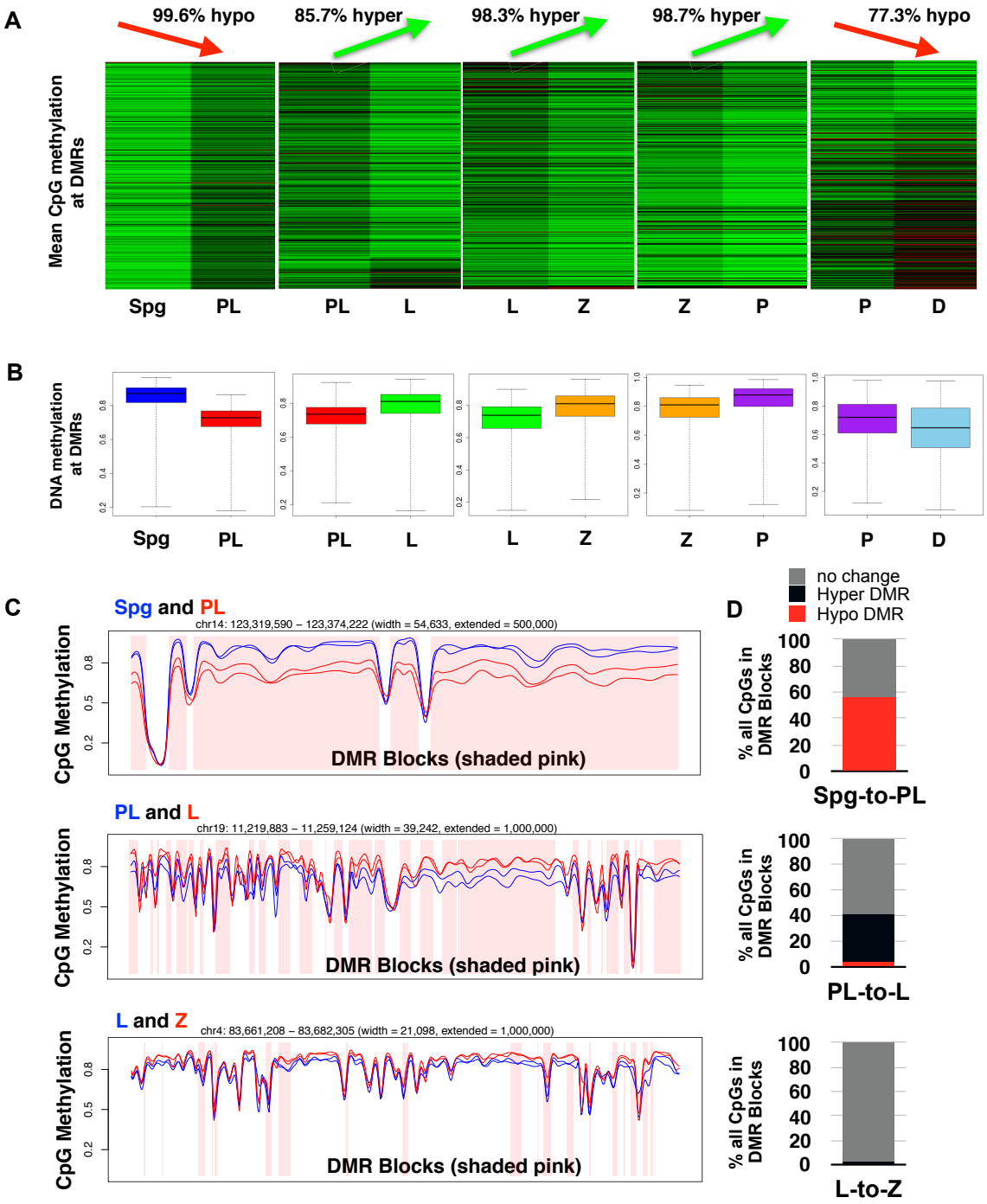

A
99.6% hypo
85.7% hyper
98.3% hyper
98.7% hyper
77.3% hypo
Mean CpG methylation at DMRs
Spg
PL
PL
L
L
Z
Z
P
P
D
B
DNA methylation at DMRs
Spg
PL
PL
L
L
Z
Z
P
P
D
C
Spg and PL
chr14: 123,319,590 – 123,374,222 (width = 54,633, extended = 500,000)
CpG Methylation
DMR Blocks (shaded pink)
PL and L
chr19: 11,219,883 – 11,259,124 (width = 39,242, extended = 1,000,000)
CpG Methylation
DMR Blocks (shaded pink)
L and Z
chr4: 83,661,208 – 83,682,305 (width = 21,098, extended = 1,000,000)
CpG Methylation
DMR Blocks (shaded pink)
D
no change
Hyper DMR
Hypo DMR
% all CpGs in DMR Blocks
Spg-to-PL
PL-to-L
L-to-Z

Fig 3-8. DMRs are found throughout MPI. (A) Heatmap showing DNA methylation profile of all individual DMRs in MPI. Each horizontal line represents DNA methylation level of a DMR for an individual group (column) of a pairwise comparison, where low-to-high DNA methylation is scaled according to red (low)-to-green (high) color scale. The main direction of change in DNA methylation, either hypomethylation (hypo) or hypermethylation (hyper) for a particular pairwise comparison is indicated above the plots with accompanying percentages of all DMRs exhibiting this change. The plot was prepared with 'gplots' R package using *heatmap.2* function, with data ordered based on mean DNA methylation difference between the groups. **(B)** Boxplot showing DNA methylation value distribution at DMRs in MPI between two consecutive stages. **(C)** Smoothed DNA methylation at DMRs for Spg and PL (top), PL and L (middle), and L and Z (bottom). The DMR blocks are shown with pink shading, were Spg and PL DMRs are largely hypomethylated with respect to PL, L is largely hypermethylated with respect to PL with DMRs appearing smaller due to remethylation. The L and Z comparison exhibits fewer DMRs, the majority of which are hypermethylated in Z. **(D)** The proportion of CpGs accounted for by hypomethylated DMRs, hypermethylated DMRs or not DMRs.

A

FEATURE	DMRs total #	CpG loci (total #)	Genomic Size (bp)	Genomic Size (Mb)	CpG loci (Median #)	DMR Size (Median, bp)	Group 1 (Median)	Group 2 (Median)	meanDiff (%)
WT Spg vs PL comparison	N/A	16,825,193							
All Filtered DMR Blocks	19019	9432366	**1374153075**	**1374.15**	257	35,223	0.8708	0.7284	**13.4%**
Hypomethylated Filtered blocks (HypoMe Spg, HyperMe PL)	68	5834	600765	0.60	90	6,980	0.3906	0.5338	-13.4%
Hypermethylated Filtered blocks (HypoMe PL, HyperMe Spg)	18951	9426532	1373552310	1373.55	258	35,440	0.8712	0.7287	13.4%
WT PL vs L comparison	N/A	16,774,643							
All Filtered DMR Blocks	22631	6927448	**1006832979**	**1006.83**	203	26,694	0.7349	0.8105	**-8.8%**
Hypomethylated Filtered blocks (HypoMe PL, HyperMe L)	19406	6222514	795569858	795.57	214	25,309	0.7123	0.8043	-9.1%
Hypermethylated Filtered blocks (HypoMe L, HyperMe PL)	3225	704934	211263121	211.26	133	39,623	0.792	0.6993	9.1%
WT L vs Z	N/A	17,641,230							
All Filtered DMR Blocks	15873	522842	**447687068**	**447.69**	130	18,711	0.7386	0.8099	**-6.9%**
Hypomethylated Filtered blocks (HypoMe L, HyperMe Z)	15605	496296	445450320	445.45	131	19,037	0.7403	0.8118	-6.9%
Hypermethylated Filtered blocks (HypoMe Z, HyperMe L)	268	26546	2236748	2.24	94.5	7,644	0.6048	0.5206	7.8%
WT Z vs P comparison	N/A	17,343,570							
All Filtered DMR Blocks	25957	9976576	**1598119443**	**1598.12**	219	27689	0.8103	0.8785	-6.6%
Hypomethylated Filtered blocks (HypoMe Z, HyperMe P)	25610	9942086	1595311046	1595.31	222	28233	0.8115	0.8798	-6.6%
Hypermethylated Filtered blocks (HypoMe P, HyperMe Z)	347	34490	2808397	2.81	98	7,318	0.6133	0.5372	7.3%
WT P vs D comparison	N/A	16,433,535							
All Filtered DMR Blocks	13955	1537911	**190147548**	**190.15**	103	11441	0.7209	0.64751	**6.9%**
Hypomethylated Filtered blocks (HypoMe P, HyperMe D)	3164	242981	34990385	34.99	67	9148	0.8161	0.8773	-5.4%
Hypermethylated Filtered blocks (HypoMe D, HyperMe P)	10791	1294930	155157163	155.16	113	12,109	0.6891	0.58794	8.2%

B

	Mean percentage (%) of genomic feature quantitation (covered by remethylating DMRs)			
Genomic Feature	**HypoMe in PL**	**HypoMe in L**	**HypoMe in Z**	**HypoMe in P**
Exons	39.81422	11.541307	38.972687	1.3070884
Introns	41.279953	11.410929	38.78446	1.3409786
CDS	34.758404	12.464094	38.96116	1.4424361
Repeats	32.167152	16.16434	57.317757	1.4923959

Table 3-7. **Results summary of Differentially Methylated regions (DMRs).** Only CpGs in common between all four samples of a pairwise comparison (two biological replicates per group) were included in the analysis. **(A)** DMRs between wild-type (WT) samples. See **Protocol** 3 (in Chapter 4) for details on computational DMR analysis. The computational comparison was performed in a temporally progressive manner, such that the difference between stages is calculated, as Group 1 (stage n) - Group 2 (stage n+1). For example, in WT Spg vs PL comparison, the difference is calculated as (Spg - PL). Note, that what is hypermethylated in Spg, is hypomethylated in PL. **(B)** Analysis of remethylating DMRs (DMRs that are still hypomethylated relative to the next stage). DMR coordinates were imported into SeqMonk program and percent of exons, introns, coding sequences (CDS) and repetitive sequences covered by DMRs was quantified. **(C)** DMRs between WT and *Mael* knock-out (KO*)* samples. The computational comparison between *Mael-/-* and WT was calculated, as Group 1 (*Mael-/-*) - Group 2 (WT). For example, in *Mael-/-* Spg vs WT Spg comparison, the difference is calculated as (mutant Spg - WT Spg).

Gene expression dynamics in MPI

In view of the presence of TRDM-dependent and independent DMRs throughout MPI, and prior evidence of correlation, at gene promoters, of hypermethylation with gene silencing, and undermethylation with gene expression [43, 157, 158], we wondered if hypomethylated DMRs associate with gene transcription. To obtain gene expression profile, we performed RNA-seq of FACS-enriched individual MPI cell populations. We used TopHat with Bowtie2 to align the reads to the mouse genome (mm9), and HTSeq package to count the reads [159, 160]. RNA-seq statistics are found in **Table 3-8**. Subsequently, we used EdgeR to obtain normalized counts per million (CPM) and perform differential expression analysis [161, 162]. For the analysis of transcript abundance, we converted normalized counts to Reads Per Kilobase of transcript per Million mapped reads (RPKM).

Analysis of RNA-seq data confirmed that our FACS-enriched populations were free of testicular somatic cells (**Fig 3-9A**) [163, 164]. Examination of gene expression characteristic of Spg confirmed enrichment of differentiated Spg (as indicated, for example, by expression of *Dmrt1*, *Dmrt6*, *Kit*) and lack of undifferentiated type A Spg or spermatogonial stem cells (based on low expression of *Plzf/Zbtb16* and *Thy1*) (**Fig 3-9B**) [120, 165]. A wide panel of genes expressed in MPI, including those involved in homologous chromosome pairing [166], meiotic initiation and meiosis-specific sister-chromatid cohesion [166, 167], DSB formation, DNA damage response and recombination [167] (**Fig 3-9C-F**), confirmed and expanded on the expression patterns of these genes during MPI. The RNA-seq experiment also confirmed that collected PL cells are highly enriched in replication-associated transcripts including replication-dependent histone genes (**Fig 3-9G-H)** [168], but not transcripts encoding isoforms of replication-independent variant H3.3 (**Fig 3-9I**) [169]. It is worth noting that although earlier studies on RNA synthesis in MPI offered invaluable insights into the many global and specific features of MPI transcription [81, 164, 170-173], most studies were performed on total juvenile testes or limited enriched MPI populations.

Sample ID	Sample Name	# Total reads	# Kept reads	# Discarded reads*	# unaligned reads	# aligned reads	% Aligned reads (aligned/total)
Sample_3	WT Spg	26,543,113	26,085,556	457,557	1,767,453	24,775,660	93.3
Sample_8	WT PL	47,294,899	46,392,718	902,181	5,138,031	42,156,868	89.1
Sample_17	WT L	40,111,575	38,395,027	1,716,548	5,021,064	35,090,511	87.5
Sample_22	WT Z	29,967,059	29,344,060	622,999	2,558,744	27,408,315	91.5
Sample_26	WT P	35,860,833	35,259,664	601,169	5,771,564	30,089,269	83.9
Sample_29	WT D	66,582,354	66,170,052	412,302	4,983,221	61,599,133	92.5
Sample_WTsperm1	WT Sperm	31,013,810	30,946,137	67,673	2,879,598	28,134,212	90.7
Sample_WTsoma1	WT Soma	45,381,435	45,295,149	86,286	3,740,655	41,640,780	91.8
Sample_KOSpg1	Mael-/- Spg1	32,337,933	32,282,394	55,539	2,470,401	29,867,532	92.4
Sample_KOSpg2	Mael-/- Spg2	31,732,462	31,635,906	96,556	4,276,568	27,455,894	86.5
Sample_KOPL1	Mael-/- PL1	36,221,878	36,099,518	122,360	4,914,404	31,307,474	86.4
Sample_31	Mael-/- PL2	44,820,294	43,342,034	1,478,260	4,739,450	40,080,844	89.4
Sample_KOPL3	Mael-/- PL3	39,508,688	39,416,155	92,533	5,104,823	34,403,865	87.1
Sample_32	Mael-/- LZp	54,405,658	54,268,517	137,141	4,413,493	49,992,165	91.9
Sample_KOsoma1	Mael-/- Soma1	59,326,582	59,258,581	68,001	2,489,858	56,836,724	95.8
Sample_KOsoma2	Mael-/- Soma2	30,328,098	30,284,214	43,884	2,210,284	28,117,814	92.7

Table 3-8. RNA-seq statistics. Alignment results to mm9 genome for 50-bp single end reads. Alignment was performed with Tophat. Note, *Tophat* discards those reads which have low complexity or high number of N. Unmapped reads were quantified from unmapped.bam using the following *samtools* <samtools view unmapped.bam | wc -l>. Aligned reads were calculated as the number of kept reads - # of unaligned reads.

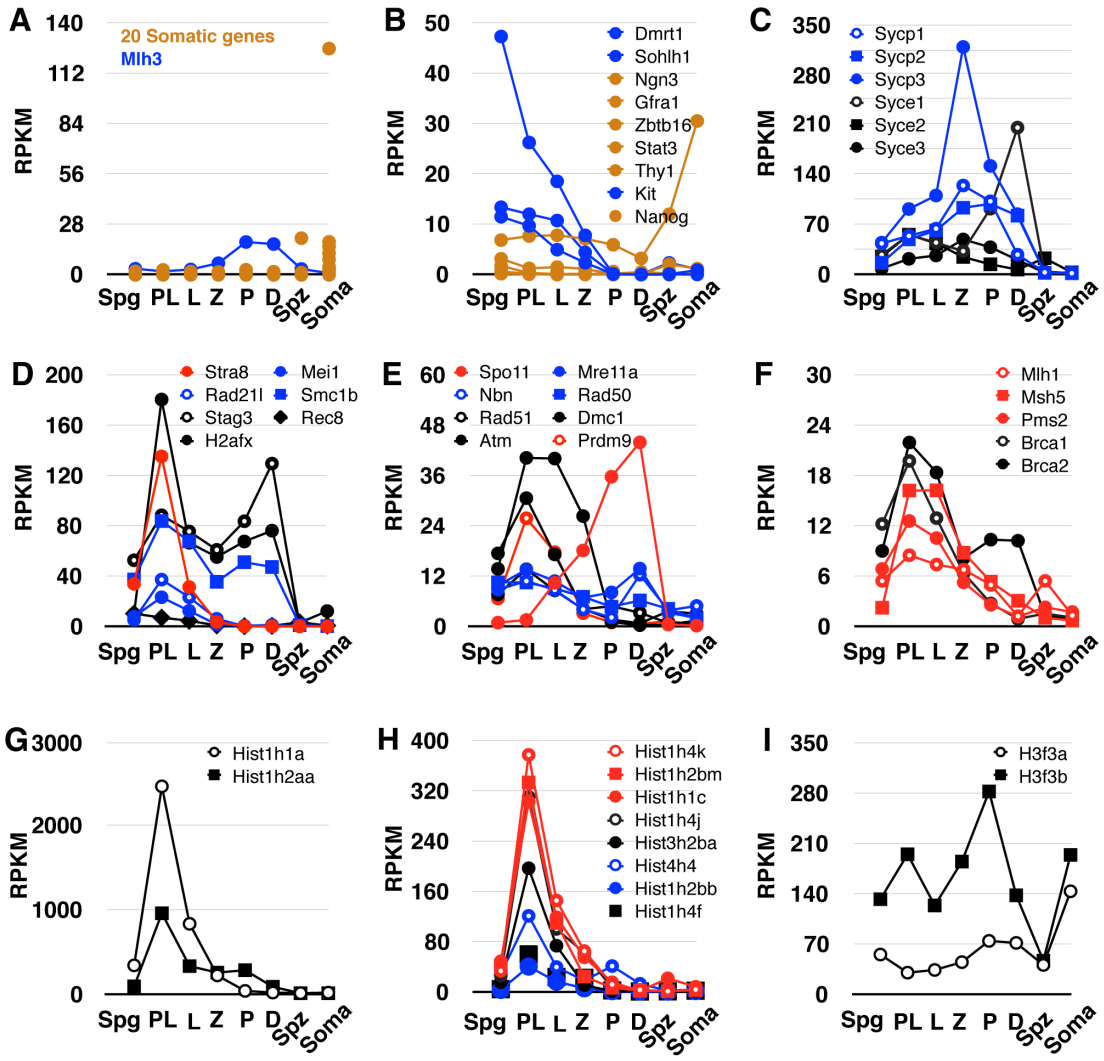

A
20 Somatic genes
Mlh3
RPKM
140
112
84
56
28
0
Spg PL L Z P D Spz Soma
B
Dmrt1
Sohlh1
Ngn3
Gfra1
Zbtb16
Stat3
Thy1
Kit
Nanog
50
40
30
20
10
C
Sycp1
Sycp2
Sycp3
Syce1
Syce2
Syce3
350
280
210
140
70
D
Stra8
Rad21l
Stag3
H2afx
Mei1
Smc1b
Rec8
200
160
120
80
40
E
Spo11
Nbn
Rad51
Atm
Mre11a
Rad50
Dmc1
Prdm9
60
48
36
24
12
F
Mlh1
Msh5
Pms2
Brca1
Brca2
30
18
6
G
Hist1h1a
Hist1h2aa
3000
2000
1000
H
Hist1h4k
Hist1h2bm
Hist1h1c
Hist1h4j
Hist3h2ba
Hist4h4
Hist1h2bb
Hist1h4f
400
320
240
160
80
I
H3f3a
H3f3b

Fig 3-9. Transcript abundance of select genes in MPI germ cells. Transcript abundance is expressed as RPKM. Prominent transcripts from **(A)** testicular somatic cells, including Sertoli, Leydig and Macrophage cells, were examined to assess the level of contamination, and included Amh, Ccl2, Cd9, Cyp11a1, Cyp17a1, Fn1, Fshr, Gap43, Gata1, Gata4, Gpc3, Lhcgr, Lum, Mmp12, Mmp9, Pla2g4a, Rlf, Star, Tead2 and Vcam1. Meiosis-specific genes Mlh3 are exhibited for relative comparison. Transcript abundance of genes associated with **(B)** differentiated and undifferentiated Spg, **(C)** meiosis-specific synaptonemal complex (SC) formation, where Sycp2 and Sycp3 encode components of the axial elements and Sycp1, Syce1-3 make up central element of the SC, **(D)** meiotic onset (Stra8), meiosis-specific sister-chromatid cohesion complex (Smc1b, Rec8, Rad21l and Stag3), recognition of meiotically-programmed DSBs (yH2AX) and other early meiosis-associated genes (e.g., Mei1), **(E-F)** DSB formation and repair and recombination, were evaluated. **(G-H)** Replication-dependent histone variant genes are highly expressed and enriched in PL spermatocytes. Twelve replication-dependent histone variant genes with high transcript abundance are shown. Selected, are whose genes that were recently shown, by quantitative RT-PCR (and in some instances *in situ* hybridization), to be highly enriched in early spermatocytes at 9-dpp testis, but not 2-dpp (gonocytes), 25-dpp (enriched in round spermatids) or 60-dpp (enriched in haploid cells) [168]. **(I)** Two genes, H3f3a and H3f3b, encoding replication-independent histone H3.3 were examined.

We first examined overall transcription dynamics in MPI. It is known that transcriptional upregulation and high levels of RNA synthesis occur in P spermatocytes and these changes are associated with reprogramming of gene expression [81, 164, 174]. Indeed, we observe that prominent changes in gene expression occur from Z to P and P to D spermatocytes (**Fig 3-10A**). However, changes in gene expression are also associated with other MPI transitions, as evident from the dynamic mean RNA expression values (**Fig 3-10B**). To identify changes in gene expression, we performed differential expression analysis between consecutive MPI substages. The significantly differentially expressed genes were defined as exhibiting (1) fold change of at least 1.5 and (2) an average pairwise expression value of at least 1 Count Per Million (CPM) and (3) false discovery rate of less than 0.05. Based on this criteria, the Spg to PL transition involved upregulation of ~2250 genes (with top genes including H2afx and other histones) and downregulation of ~3475 genes. Interestingly, during PL stage there is an increase in transcript abundance of many genes unique to MPI (**Fig 3-9D-F**). The PL to L transition involved upregulation of ~1700 genes (top genes included *Spo11* and *Meiob*), and downregulation of ~1730 genes (top genes included *Stra8*). The L to Z transition was associated with upregulation of ~2240 genes (including meiotic DSB repair genes *Msh4*, *Hormad1* and *Hormad2*) and downregulation of ~2990 genes (top genes including *Stra8*, *Mei1* and *Prdm9* which were upregulated earlier in PL). The transition from Z to P involved upregulation of ~3900 genes (many related to sperm function), and downregulation of ~3680 genes (among top genes are X-linked genes). P to D transition involved upregulation of ~3600 genes (including many zinc finger genes, and genes associated with protein modifications, like ubiquitination), and downregulation of ~3200 genes (many associated with nucleic acid metabolism).

We evaluated transcript abundance of genes associated with passive or active DNA demethylation. Our RNA-seq data is consistent with an idea of passive replication-dependent DNA demethylation in PL, but does not support a significant role for Tet-mediated active DNA demethylation, due to the absence of *Tet1* and *Tet2* and very low transcript abundance of *Tet3*

(**Fig 3-10A**). However, locus-specific TET-mediated DNA demethylation could not be excluded and 5-hydroxymethylcytosine (5hmC), a product of TET-mediated oxidation of 5mC, is present in PL and early MPI cells [175]. PL cells also seem to lack transcripts for genes involved in other means of active DNA-demethylation, including Activation-Induced Cytidine Deaminase (*Aicda/Aid*), Apolipoprotein B MRNA Editing Enzyme (*Apobec1*), Thymine DNA Glycosylase (*Tdg*) and Single-Strand-Selective Monofunctional Uracil-DNA Glycosylase 1 (*Smug1*) (**Fig 3-11A Fig**).

We also evaluated transcript abundance of genes associated with maintenance or *de novo* DNA remethylation. PL is enriched for DNMT1, a maintenance methyltransferase that associates with replication foci and remethylates hemimethylated DNA generated during replication (**Fig 3-11A**). [176, 177] and contains *Uhrf1*, which is essential for targeting DNMT1 to chromatin during S phase [176, 178]. We observe low abundance of transcripts associated with *de novo* DNA methylation, including DNMT3A and DNMT3B, and very low abundance of Dnmt3l transcripts (**Fig 3-11A**). These observations are consistent with previously reported down-regulation of DNMT3A and DNMT3l based on RT-qPCR and protein localization studies in juvenile testis [56]. However, low transcript abundance does not rule out the contribution of *de novo* DNA methylation to remethylaton, and indeed, protein levels of DNMT3A and isoform DNMT3A2 are detectable by immunofluorescence, although barely detectable for DNMT3B in EdU-positive PL cells (**Fig 3-11B**). However, it is important to note that transcripts for Dnmt3l are nearly absent in PL (~0.5 RPKM), and DNMT3l stimulates de novo DNA methylation by DNMT3A and DNMT3B [179]. Our RNA-seq data supports passive replication-driven DNA demethylation, and remethylation of DNA primarily by a maintenance-related mechanism, and argues against major contribution of *de novo* DNA methylation.

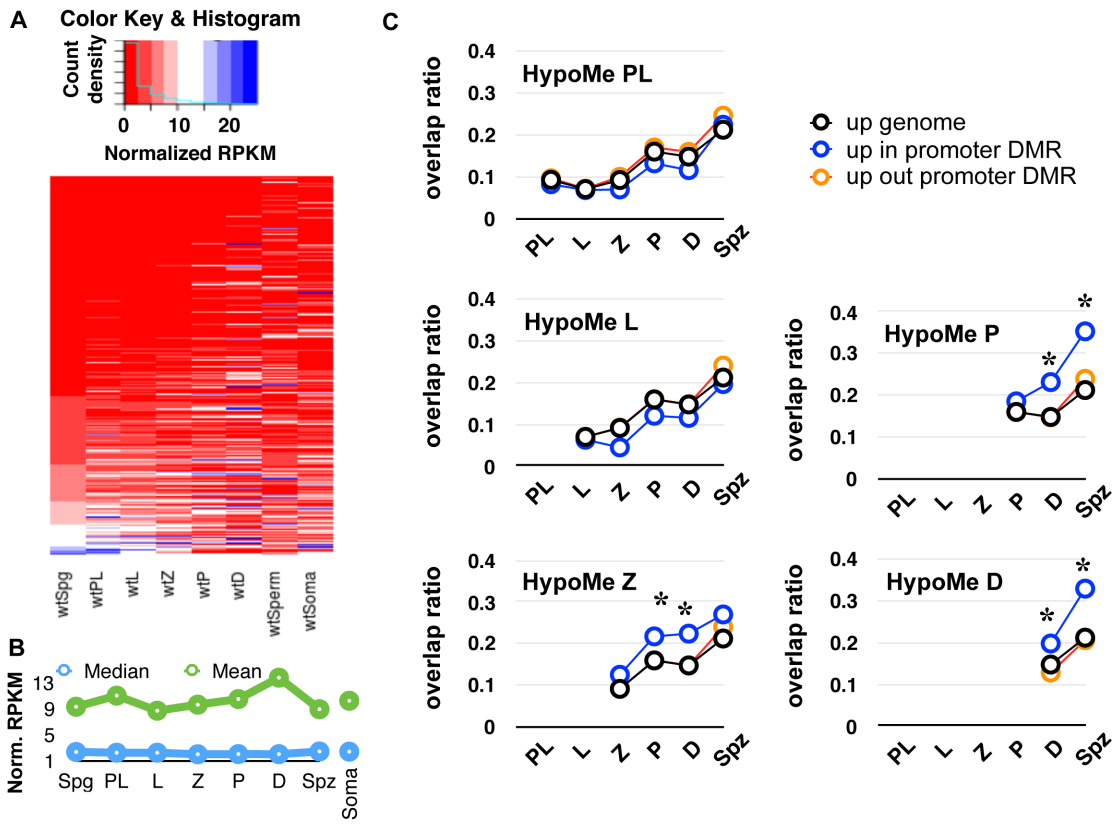


Fig 3-10. DMRs found in mid- and late- MPI are correlated with gene transcription in late MPI and Spz. **(A)** Heatmap showing clustering of gene expression in Spg, MPI, sperm and soma. Shown are expressed genes with RPKM values of < 25 (genes with RPKM<1 in all eight samples were eliminated from this analysis), accounting for 80% all expressed genes throughout MPI (12128 genes). Genes with >25 RPKM were excluded from this analysis to allow simultaneous visualization of the majority of genes for all the samples using "heatmap.2" function from 'gplots' package in R. Data was sorted based on ascending RPKM values in Spg sample. **(B)** The mean and median normalized RPKM values for all expressed genes (n=15170) is shown. **(C)** Enrichment of overlap between promoters within hypomethylated DMRs and upregulated genes was measured by overlap ratio analysis. 'Up genome' denotes ratio of all upregulated genes at a particular MPI stage or Spz relative to all genes expressed in the genome at the time; 'up in promoter DMR' denotes ratio of upregulated genes inside promoter DMRs, relative to all genes found inside the DMRs; 'up out promoter DMRs' indicates a ratio for upregulated genes outside

of promoter DMRs, relative to all genes found outside of DMRs. The asterisk (*) indicates a p-value of < 0.05 as calculated using Fisher's exact test. For each pairwise comparison, the 2x2 table containing the (a) number of up genes inside the promoter DMRs (b) number of genes outside the promoter DMRs, (c) number of other genes inside and (d) number of other genes outside, was formed and Fisher's exact test performed.

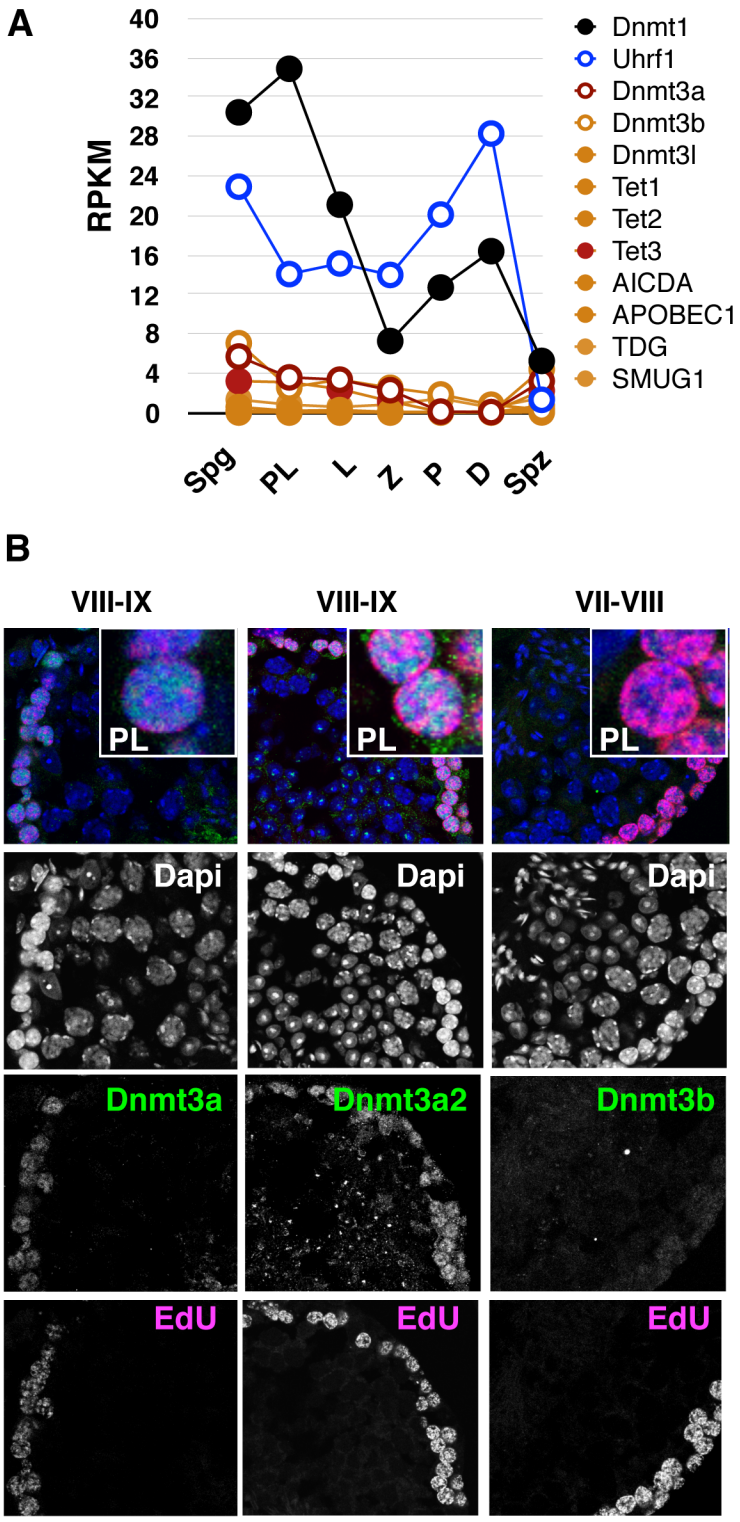

A
RPKM
40
36
32
28
24
20
16
12
8
4
0
Spg
PL
L
Z
P
D
Spz
Dnmt1
Uhrf1
Dnmt3a
Dnmt3b
Dnmt3l
Tet1
Tet2
Tet3
AICDA
APOBEC1
TDG
SMUG1
B
VIII-IX
VIII-IX
VII-VIII
PL
Dapi
Dnmt3a
Dnmt3a2
Dnmt3b
EdU

Fig 3-11. Examination of transcript abundance and proteins associated with passive or active DNA demethylation or remethylation. (A) Transcript abundance of genes associated with passive or active DNA demethylation. **(B)** Immunolfuorescence of proteins associated with de novo DNA methylation, Dnmt3a, Dnmt3a2 and Dnmt3b, in the context of EdU staining for replicative DNA and DNA counterstaining with DAPI.

The transcriptional activity of sex chromosomes becomes suppressed during male meiosis due to epigenetic chromatin modifications associated with meiotic sex chromosome inactivation, or MSCI [30]. Accordingly, global reduction of expression levels of protein-coding genes on chrX has been observed in P spermatocytes [81, 172, 173]. Indeed, our analysis shows sharp reduction of gene expression on chrX in P spermatocytes and maintenance of low transcription in D spermatocytes (**Fig 3-12A,B**), however, consistent with previous reports, transcriptional reactivation of some genes is observed in D [81]. Pairwise comparison of transcript abundance between consecutive MPI stages reveals that although gene transcription on chrX is strongly correlated between Spg and PL, PL and L, and L and Z, there is a decrease in transcript abundance during L to Z transition as well, and this is not a general phenomenon, since chr1 exhibits upregulation of transcription at this time instead (**Fig 3-12A**). Indeed, mean transcription levels for chrX reveal that there is a gradual decrease in gene transcript abundance on chrX as early as PL to L transition (**Fig 3-12B**), with highest levels of expression observed in pre-meiotic Spg, epididymal Spz and testicular somatic cells (**Fig 3-12B**). In view of our observed gradual remethylation dynamics on chrX (and genome-wide) (**Fig 3-4**) and gradual shutdown of transcription on chrX (**Fig 3-12A,B**), a progressive gain in DNA methylation after TRDM in MPI is coincident or may be correlated with the establishment of MSCI.

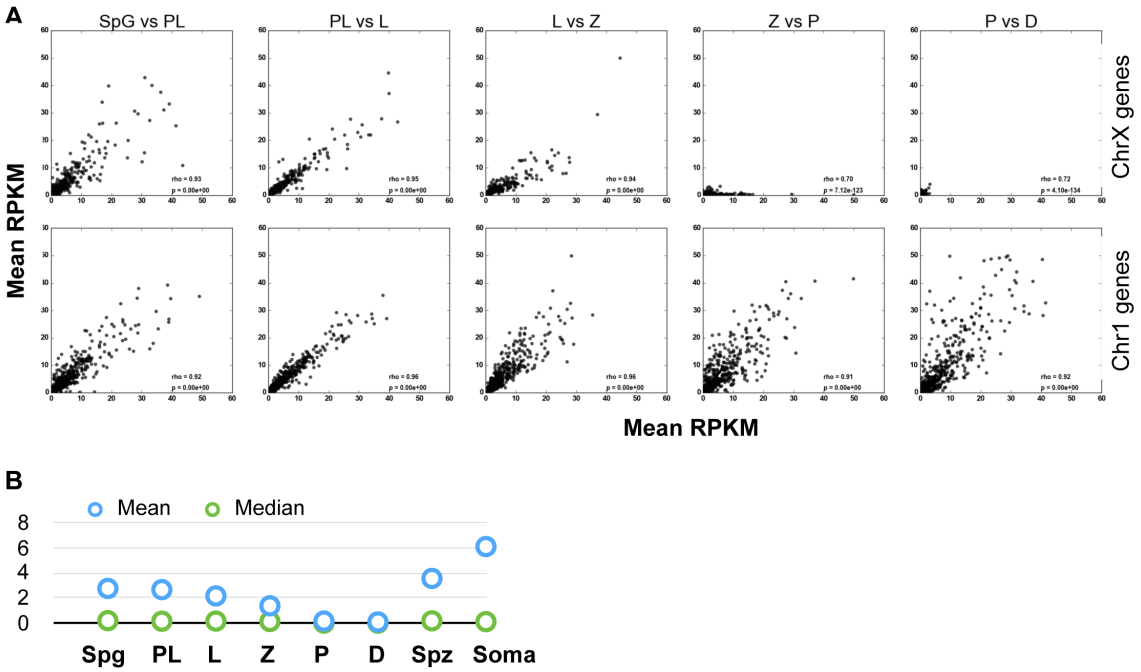


Fig 3-12. Transcript abundance of chromosome X (chrX)-linked and chromosome 1 (chr1)-linked genes. (A) Mean transcript abundance (RPKM) was examined for meiotic prophase I stages and pre-meiotic Spg in a pairwise manner for genes on chrX or chr1. For best visual analysis, the majority of genes, with RPKM values of less than 50, were evaluated, while the outliers were removed. Spearman's rank correlation coefficient (rho), and p-value were computed for each pairwise analysis. **(B)** Mean and Median RPKM values (y-axis) for all genes expressed from chromosome X.

Differentially methylated regions are associated with gene transcription.

Although the relationship between DNA methylation and transcription is complex, promoter DNA methylation can inhibit transcription, directly or indirectly (e.g., via recruitment of repressive complexes and histone deacetylases) [42, 43, 157]. Similarly, although many CpG-rich promoters are unmethylated in all cell types independent of the state of gene expression, promoter hypomethylation is associated with transcriptional activation in developmental and disease contexts [180-182]. To determine the extent of association of gene expression with DNA hypomethylation in MPI, we identified those promoters that were enriched by at least 90% of their length in hypomethylated DMRs. Next, we evaluated the degree of overlap between these hypomethylated promoter DMRs and genes upregulated during MPI (**Table 3-9C, Methods and Procedures**). The analysis suggested that an overlap between promoter hypomethylation and gene upregulation is not random for DMRs that become hypomethylated in Z, P and D, since the overlap ratio was higher for upregulated genes inside the promoter DMRs than outside the promoter DMRs and genome-wide (**Fig 3-10C, Table 3-9**). To test the significance of the enrichment of hypomethylated DMRs with upregulated genes we performed Fisher's exact test (**Methods and Procedures**) (**Fig 3-10C,** significance marked by asterisk, **Table 3-9**). The statistical analysis, which was mindful of upregulated genes inside the promoter DMRs, upregulated genes outside of promoter DMRs and all other genes inside or outside promoter DMRs, revealed a significant association between a small subset of genes that becomes hypomethylated in Z and transcriptionally upregulated in P, D and Spz. Similarly, a significant association existed between a subset of genes that becomes hypomethylated in P and transcriptionally upregulated in D and Spz, while hypomethylation of more than a thousand promoters in D was associated with upregulation of these genes in D and Spz (**Fig 3-10C**). Importantly, PL and L stages were not associated with transcriptional upregulation, suggesting that hypomethylation in early MPI is independent from transcriptional upregulation. While relationship between gene expression and DNA methylation is complex, and unmethylated

promoter CpGs are not necessarily associated with active transcription (especially constitutively unmethylated regions) [183]. Our results indicate that there is a significant association between coding-gene promoter hypomethylation in Z, P and D and gene expression in P, D and Spz.

	Hypo PL up PL	Hypo PL up L	Hypo PL up Z	Hypo PL up P	Hypo PL up D	Hypo PL up Spz
up	2239	1701	2226	3892	3598	5172
total	24395	24395	24395	24395	24395	24395
up/total	**0.0917811**	**0.0697274**	**0.0912482**	**0.1595409**	**0.1474892**	**0.2120107**
up in	516	427	437	834	731	1422
total in	6351	6351	6351	6351	6351	6351
up in/total in	**0.081247**	**0.0672335**	**0.0688081**	**0.1313179**	**0.1151**	**0.2239017**
up out	1723	1274	1789	3058	2867	4435
total out	18044	18044	18044	18044	18044	18044
up out/total out	**0.0954888**	**0.0706052**	**0.0991465**	**0.1694746**	**0.1588894**	**0.2457881**

	Hypo L up L	Hypo L up Z	Hypo L up P	Hypo L up D	Hypo L up Spz
up	1701	2226	3892	3598	5172
total	24395	24395	24395	24395	24395
up/total	**0.0697274**	**0.0912482**	**0.1595409**	**0.1474892**	**0.2120107**
up in	24	17	46	44	75
total in	382	382	382	382	382
up in/total in	**0.0628272**	**0.0445026**	**0.1204188**	**0.1151832**	**0.1963351**
up out	1677	2209	3846	3554	5782
total out	24013	24013	24013	24013	24013
up out/total out	**0.0698372**	**0.0919918**	**0.1601632**	**0.1480032**	**0.2407862**

	Hypo Z up Z	Hypo Z up P	Hypo Z up D	Hypo Z up Spz
up	2226	3892	3598	5172
total	24395	24395	24395	24395
up/total	**0.0912482**	**0.1595409**	**0.1474892**	**0.2120107**
up in	17	33	34	41
total in	136	152	152	152
up in/total in	**0.125**	**0.2171053**	**0.2236842**	**0.2697368**
up out	2209	3859	3564	5816
total out	24243	24243	24243	24243
up out/total out	**0.0911191**	**0.15918**	**0.1470115**	**0.2399043**

	Hypo P up P	Hypo P up D	Hypo P up Spz
up	3892	3598	5172
total	24395	24395	24395
up/total	**0.1595409**	**0.1474892**	**0.2120107**
up in	44	55	84
total in	238	238	238
up in/total in	**0.1848739**	**0.2310924**	**0.3529412**
up out	3848	3543	5773
total out	24157	24157	24157
up out/total out	**0.1592913**	**0.1466656**	**0.2389783**

	Hypo D up D	Hypo D up Spz
up	3598	5172
total	24395	24395
up/total	**0.1474892**	**0.2120107**
up in	1333	2220
total in	6735	6735
up in/total in	**0.1979213**	**0.3296214**
up out	2265	3637
total out	17660	17660
up out/total out	**0.1282559**	**0.2059456**

Table 3-9. **Enrichment of overlap between DMRs and gene expression.** We measured an overlap ratio between gene promoters inside hypomethylated DMRs and transcriptionally upregulated genes. Fisher's exact test was performed.

Wild-type meiotic onset is accompanied by transposon expression, which if not silenced results in TE reactivation and meiotic demise

Landmark reactivation of TEs in MPI is observed in various mutants defective in DNA methylation and piRNA pathway mutants [77, 94, 116, 140, 141]. These mutants, including *Dnmt3*l-/-, *Miwi2-/-* or *Mael-/-*, exhibit a complete block of spermatogenesis due to defects in homologous chromosome synapsis, DNA damage, derepression of L1 retrotransposons and apoptosis in MPI [77, 116, 141]. These mutants seem to exhibit massive TE reactivation specifically in MPI, and not before [77, 116]. Thus, it has been suggested that meiosis is a "weak link" when it comes to TE control. The reasons for this TE reactivation pattern in the mutants remain poorly understood, although recent evidence suggests that in the absence of DNA methylation (in the case of *Dnmt3l* mutant) premature loss of H3K9me2 triggers TE re-activation [77]. We thus wondered if DNA demethylation at the onset of meiosis in PL contributes to TE expression, and thereby creates an opportunity for massive TE activation in those mutants that are defective in transposon silencing. Thus, we hypothesized, that a window of relaxation of TE control in wild type MPI, via TRDM, may serve as a developmental window of opportunity for TE de-repression in various DNA methylation and piRNA pathway mutants exhibiting abberant TE upregulation in MPI. Thus, our data on L1Orf1p expression suggests that it may not be that the mutants create an opportunity for TE expression, but that the opportunity is already there in the WT MPI, and the mutants are simply taking advantage of it.

We investigated *Mael-/-* and performed WGBS-seq and RNA-seq analysis on FACS-enriched Spg, PL and a combined spermatocyte population we call "LZp" [146]. The LZp population was collected as a single population due to the difficulty to separate normal L and Z spermatocytes from abnormal cells with mixed L and Z characteristics and a minor population of abnormal P-like spermatocytes [116]. WGBS-seq on FACS-enriched spermatocytes from *Mael* mutant yielded comparable Illumina sequencing read number, quality, alignment and bisulfite-conversion rate statistics as WT samples (**Table 3-10**). We find that similarly to WT, *Mael-/-*

exhibits TRDM, followed by DNA remethylation (**Table 3-2**), suggesting that regain of DNA methylation at most genomic sequences does not depend on MAEL expression. Nevertheless, the overall levels of DNA methylation in pre-meiotic Spg were lower than in the WT, suggesting the inheritance of low DNA methylation from Spg precursors. We speculate that DNA methylation defects in Spg precursors arise after birth and are uncoupled from fetal defects in DNA methylation, since it was previously shown that in the *Mael* mutant DNA methylation at L1 recovers to WT-like levels by birth [127].

Sample ID	Sample Name	# Raw reads	# Failed to align reads	% overall read alignment	# Raw read pairs	# uniquely-aligned read pairs**	% Uniquely aligned read pairs	# Non-uniquely aligned read pairs	# unique alignment pairs after de-duplication
VG15	KO Spg1	426,905,604	25,069,600	94.1%	213,452,802	168,069,776	78.7%	20,313,426	72,568,802 (43.18% of total)
VG17	KO PL1	273,274,082	14,449,488	94.7%	136,637,041	111,743,750	81.8%	10,443,803	36,446,773 (32.62% of total)
VG18	KO PL2	431,529,226	22,849,626	94.7%	215,764,613	173,186,881	80.3%	19,728,106	65,091,226 (37.58% of total)
VG16	KO PL 3	411,349,320	22,146,388	94.6%	205,674,660	164,764,341	80.1%	18,763,931	64,482,493 (39.14% of total)
VG19	KO LZP-1	357,041,454	33,574,191	90.6%	178,520,727	129,554,013	72.6%	15,392,523	35,124,935 (27.11% of total)
VG21	KO LZP-2	426,170,156	23,123,249	94.6%	213,085,078	168,605,295	79.1%	21,356,534	57,633,993 (34.18% of total)
VG24	WT Sperm1	148,205,076	7,576,682	94.9%	74,102,538	59,926,611	80.9%	6,599,245	30,188,571 (50.38% of total)
VG25	WT Sperm2	131,808,906	7,211,961	94.5%	65,904,453	52,411,892	79.5%	6,280,600	34,078,847 (65.02% of total)

Table 3-10. **Whole-Genome Bisulfite Illumina (WGBS) sequencing results.** Sequencing read alignment and de-duplication summary for *Mael* knock-out (KO*)* samples are shown. A read is said to align "uniquely" if it has exactly one valid alignment to the reference according to the alignment policy. Percent (%) overall alignment was calculated as (# of raw reads - # of failed to align reads)/total # reads.

We observe that *Mael* mutant maintains relatively high DNA methylation levels at major classes of TEs, namely, the LINE (**Fig 3-13A**), SINE, LTR and DNA (**Fig 3-14**), despite the lower starting levels in mutant Spg compared to WT Spg. However, we find that in *Mael-/-*, the young L1 families with intact and potentially active members, including some of the youngest members like L1Md_Gf, exhibit much lower levels of DNA methylation from the start (**Fig 3-13B**). This is in contrast to the relatively higher levels of DNA methylation at older families of LINEs, like Lx8 (**Fig 3-13B)**. These results suggest an important role for MAEL in piRNA-mediated DNA methylation of younger families of L1.

We asked if TRDM of potentially active LINEs in PL contributes to their transcription in WT and *Mael-/-* mutants. First we evaluated overall gene transcript abundance in *Mael-/-* compared to WT using principal component analysis of normalized RNA counts (**Fig 3-15**). This analysis revealed that transcriptomes for Spg, PL and spermatocytes, but not testicular somatic cells, cluster separately and thus differ. Indeed, while transcript abundance for most genes was similar between the WT (**Fig 3-15B, solid lines**) and *Mael-/-* **(Fig 3-15B, dashed lines**), there were many genes that appeared changed. Gene enrichment analysis on significantly downregulated genes revealed that over a thousand genes were downregulated in *Mael-/-* PL and these were related to processes of spermatogenesis, gamete generation and chromosome organization. Several thousand genes also appeared upregulated in *Mael-/-* and these genes were most significantly associated with defense responses including immune response and response to stress.

To analyze RNA abundance of TE we used RepEnrich strategy in order to account for most TE-derived RNA in our data, by way of counting both, uniquely mapped and multi-mapped reads [184]. In the WT, we find that transcript abundance for repeat elements as a whole shows an overall decrease from Spg onwards, with lowest levels in Spz (**Fig 3-13C,** green). This is in contast to *Mael-/-*, which exhibits gradual upregulation of TEs from Spg onwards, until the highest levels in meioitic LZp population (**Fig 3-13C,** magenta). The testicular somatic cells in

both, WT and *Mael-/-,* exhibited similar levels of TE abundance. Intriguingly, we find that in WT, Spg to PL and PL to L transitions are accompanied by transcriptional upregulation of many class LINE elements **(Fig 3-13D,** top**)**. This upregulation includes all classes of potentially active L1 elements **(Fig 3-13D,** bottom**)**, whose expression begins to decrease in Z and is essentially extinguished by P (**Fig 3-13D, bottom**). Intriguingly, in the WT, P-to-D transition involves a dramatic upregulation of LINEs and potentially active LINE members (**Fig 3-13D**). Our findings provide evidence that in WT MPI, TE transcription occurs in PL and increases through L stages, demonstrating that meiotic onset contributes to TE expression. In view of our DNA methylation analysis, it is plausible that TRDM in PL and L contribute to this transposon expression. In view of this finding, it is interesting that TE upregulation in the *Mael-/-* also commences in PL, suggesting that *Mael-/-* takes advantage of the natural TE derepression at meiotic onset (**Fig 3-13E**). However, unlike in the WT, *Mael-/-* bypasses TE silencing of young, potentially active LINE-1 elements with this subset of TEs further increasing in transcript abundance from PL to LZp (**Fig 3-13F**). This is not the case for the majority of transcripts derived from old, inactive LINE sequences, which after their initial upregulation in PL are downregulated later in MPI, in cells of the LZp population (**Fig 3-13G**). LINE1 subfamilies particularly enriched in the young and active members, namely, the L1Md_Gf, L1Md_A and L1Md_T, are among the top upregulated retrotransposons in the mutant, emphasizing element-specific defects. Our data suggest that pre-meiotic replication in PL contributes to derepression of TE silencing in *Mael-/-* even prior to meiotic entry and that silencing of TEs during TRDM is critical.

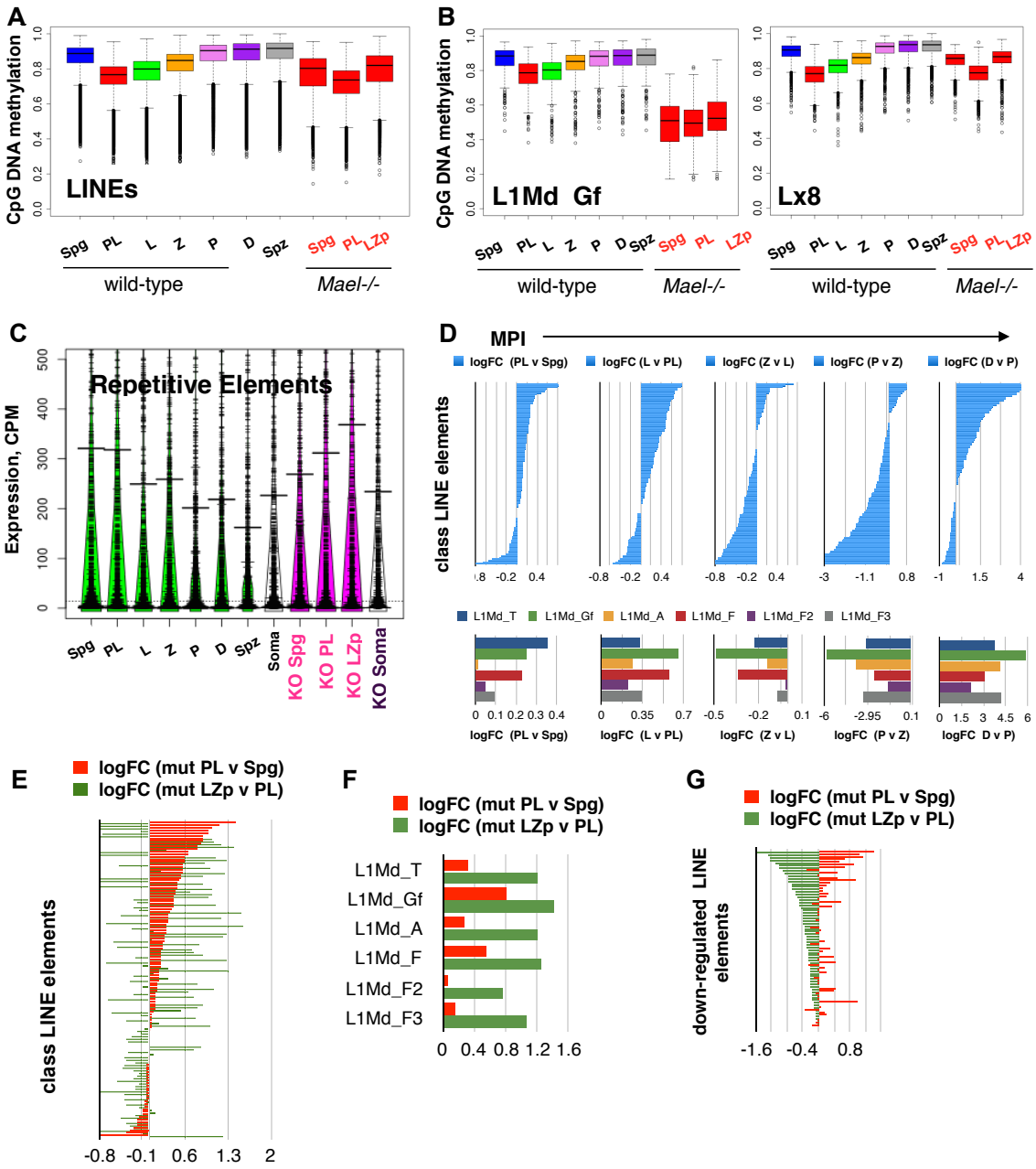
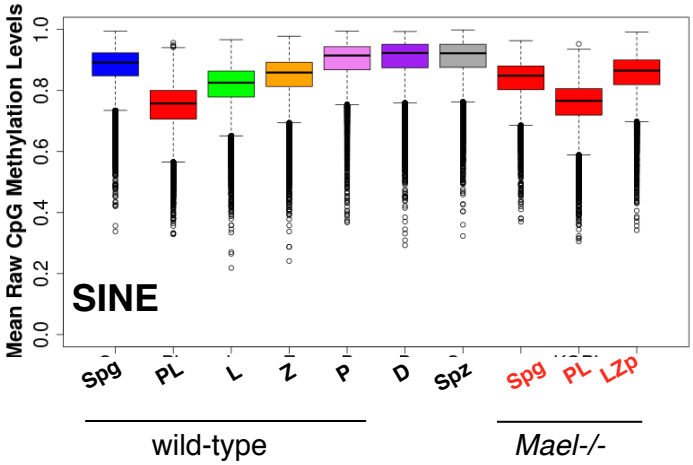
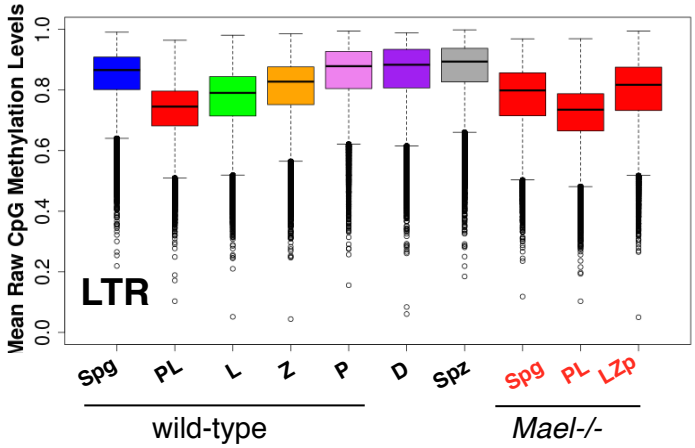

A
CpG DNA methylation
LINEs
Spg PL L Z P D Spz Spg PL LZp
wild-type
Mael-/-
B
CpG DNA methylation
L1Md Gf
Lx8
Spg PL L Z P D Spz Spg PL LZp
wild-type
Mael-/-
C
Expression, CPM
Repetitive Elements
Spg PL L Z P D Spz Soma KO Spg KO PL KO LZp KO Soma
D
MPI
logFC (PL v Spg)
logFC (L v PL)
logFC (Z v L)
logFC (P v Z)
logFC (D v P)
class LINE elements
L1Md_T
L1Md_Gf
L1Md_A
L1Md_F
L1Md_F2
L1Md_F3
E
logFC (mut PL v Spg)
logFC (mut LZp v PL)
class LINE elements
F
logFC (mut PL v Spg)
logFC (mut LZp v PL)
L1Md_T
L1Md_Gf
L1Md_A
L1Md_F
L1Md_F2
L1Md_F3
G
logFC (mut PL v Spg)
logFC (mut LZp v PL)
down-regulated LINE elements

Fig 3-13. Dynamics of MPI DNA methylation and transcript abundance in wild-type and *Mael-/-* testis. (A) DNA methylation profile of class LINE retrotransposons in WT and *Mael-/-*. WT samples include Spg (blue), PL (red), L (green), Z (orange), P (violet), D (purple), Spz (grey). *Mael-/-* samples include Spg, PL and LZp in red color. DNA methylation was analyzed as averages of consecutive, non-overlapping bins of 100 CpGs. **(B)** Examples of DNA methylation profiles corresponding to two different LINE types are shown. One, L1Md_Gf, is a young family, containing members with full-length, potentially active elements. Another, Lx8, contains ancient and inactive members. **(C)** Analysis of transcript abundance of repetitive elements for wild-type and Mael-/- after processing with RepEnrich. A beanplot was prepared with 'beanplot' R package. **(D)** A pairwise differential expression analysis is represented as fold change in normalized transposon counts (counts per million) between WT MPI stages. The horizontal barplot shows log2(FC) on the x-axis and a different LINE transposon from Repeat masker, on the y-axis. Top panels depict all class LINE elements, while bottom panels depict select LINE-1 families that contain young and potentially active members (L1Md_T, L1Md_Gf and L1Md_A) and their progenitors (L1Md_F, L1Md_F2 and L1Md_F3). The differential expression analysis was performed in EdgeR. **(E)** Differential expression analysis (represented as in (D)) is shown for *Mael-/-*. The log2(FC) of normalized transposon counts is shown for mutant PL compared to mutant Spg (red), and mutant LZp compared to mutant PL (green), where the same level red and green bars represent a particular class LINE element. **(F)** The log2 (FC) of normalized transposon counts for select LINE-1 members. **(G)** The log2(FC) of normalized transposon counts for those LINE elements in *Mael-/-* that exhibit upregulation during PL to Spg transition (red), but downregulation (green) in LZp, contrary to the potentially active LINE-1 elements that exhibit further upregulated in LZp, as shown in Fig 6F (green).


Mean Raw CpG Methylation Levels
0.0
0.2
0.4
0.6
0.8
1.0
SINE
Spg
PL
L
Z
P
D
Spz
Spg
PL
LZp
wild-type
Mael-/-


Mean Raw CpG Methylation Levels
0.0
0.2
0.4
0.6
0.8
1.0
LTR
Spg
PL
L
Z
P
D
Spz
Spg
PL
LZp
wild-type
Mael-/-

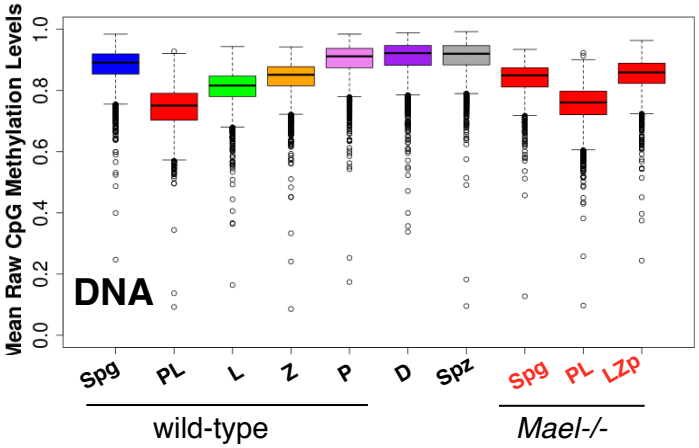

Mean Raw CpG Methylation Levels
0.0
0.2
0.4
0.6
0.8
1.0
DNA
Spg
PL
L
Z
P
D
Spz
Spg
PL
LZp
wild-type
Mael-/-

Fig 3-14. DNA methylation dynamics of major classes of transposable elements in wild type (WT) and *Mael-/-* MPI. DNA methylation levels from both replicates were pooled and averaged in consecutive, non-overlapping bins of 100 CpGs. CpGs from LINE, SINE, LTR and DNA transposons were examined. Mutant MPI exhibits similar dynamics as the WT in terms of hypomethylation followed by remethylation, but mutant Spg starts with somewhat lower DNA methylation levels. WT samples include Spg (blue), PL (red), L (green), Z (orange), P (violet), D (purple), Spz (grey). Mael -/- samples include Spg, PL and LZp in red color.

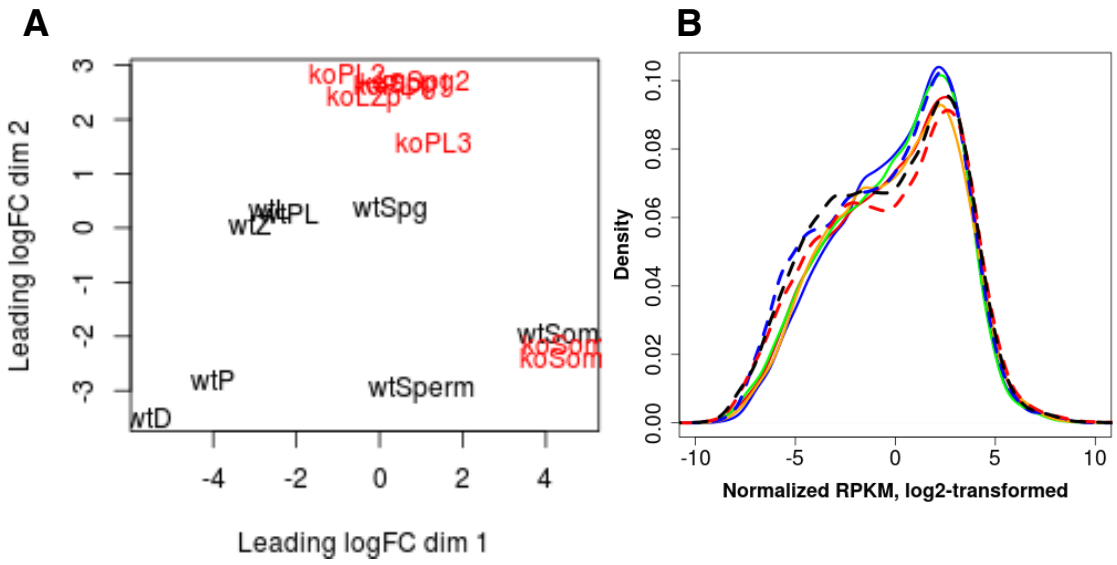


Fig 3-15. Transcriptome comparison between wild-type and Mael-/- MPI. RNA-seq reads were processed with RepEnrich pipeline to obtain transcriptional landscape of repetitive elements in wild-type (WT) and Mael-/- germ cells. (A) MDS plot of normalized counts. WT samples are in black, and mutant in red. (B) Density plot of transcript abundance (log2(FPKM)) for WT (solid lines) and Mael-/- (dashed lines).

Discussion

In this study we provided evidence for genome-wide transient relaxation of DNA methylation (or TRDM) at the onset of meiosis in adult male mice. Our findings are most consistent with TRDM by passive replication-coupled DNA demethylation in pre-meiotic S phase. This conclusion is based on a genome-wide drop of DNA methylation levels and low sequencing read coverage at early replicating domains in PL, a measure of underreplication. We also observe a hypomethylated (but DNA replicated) state of late replicating domains in L, the first stage of MPI. This is an unexpected finding since maintenance DNA methylation occurs within 2 minutes following the incorporation of cytosine in the new DNA strand during mitotic replication [185, 186]. Additionally, robust maintenance of DNA methylation in S-phase is further supported by an observation of similar levels of DNA methylation in G0/G1 and G2/M phases in primary dermal fibroblasts [187]. Thus, hypomethylation at late replicating domains in L may suggest a delay in maintenance DNA methylation in pre-meiotic S-phase. Another possibility, is that having fewer replication origins in meiotic S phase [17, 20], may contribut to the late replicating regions undergoing replication so late that their full remethylation only occurs in L or PL-to-L transition. Additional possibility is substrate limitation, namely, meiotic S phase cells being constrained in the number of methyl groups or methyl group donors. Interestingly, limitation of S-adenosylmethionine (primary methyl group donor) was found to induce cell cycle arrest in G1 and block entry into S phase [188]. Additionally, methionine depletion reduced progression rate through S phase but did not block DNA replication of cells that commit to S phase [188]. Thus, it is plausible that limiting levels of methyl groups could contribute to prolonged meiotic S phase. It is also plausible that some other gene products, important for timely remethylation, are absent or limiting, during pre-meiotic S phase.

Our data do not support a significant role for active DNA demethylation in PL due to the lack of *Tet1*, *Tet2*, *Aicda*/*Aid*, *Apobec1*, *Tgd* and *Smug1* expression, and very low *Tet3* transcript abundance. However, locus-specific Tet-mediated DNA demethylation could not be excluded and

5-hydroxymethylcytosine (5hmC), a product of Tet-mediated oxidation of 5mC, is present in PL and early MPI cells [175]. Cumulatively, these data suggest that DNA demethylation during transient relaxation of DNA methylation in PL is a result of passive replication-dependent mechanism.

Our study suggests that reacquisition of DNA methylation after hypomethylation in pre-meiotic S phase is a gradual and uneven process that creates initially large hypomethylated domains that disappear by P when the genome attains pre-meiotic DNA methylation levels. Our data suggest that DNA methylation returns gradually to pre-meiotic levels over the period of up to 70 hours (the approximate time it takes to proceed through L and Z stages of MPI) [13]. As of now, the dynamics of remethylation after TRDM cannot be explained by their preferential enrichment for specific genomic features. In addition to TRDM-dependent DMRs, we have also uncovered narrow domains of apparent TRDM-independent hypomethylation between P and D. Our analysis suggests that many of these hypomethylated domains overlap promoters of genes transcribed at P, D and later in spermatogenesis and likely arise by active DNA demethylation.

We have previously proposed that transient relaxation of transposon silencing in MPI indicated remodeling of meiotic chromatin [134]. However, it remained unknown whether the phenomenon is limited to transposons or is a reflection of a broader, genome-wide process. This work establishes that transient relaxation of transposon silencing is likely the product of a global change in DNA methylation levels of meiotic chromosomes. Interestingly, our RNA-seq experiments revealed two periods of increased expression of LINE RNA – first at the time of TRDM and second at P-to-D transition. Despite lower LINE RNA expression levels during the former, L1ORF1p expression is readily detected in early MPI. In contrast, a sharp increase in LINE RNA levels in the P-D window does not lead to the appearance of L1ORF1p. Absence of detectable levels of L1ORF1p at this time may be suggestive of efficient translational control of LINE RNAs. Intriguingly, the upregulation of LINE RNA we observe in P to D transition, also corresponds to activation of piRNA surveillance [189]. It is thus plausible that a burst in

retrotransposon transcription during P-to-D transition is involved in nucleating piRNA-mediated immunity at this time, instead of making ORF1p protein. Additionally, during late P, a large number of genes are transcribed but remain translationally repressed until many days later, when translation resumes in elongating spermatids [190]. It is plausible, that at this time uncoupling of transcription and translation occurs for LINE transcripts as well, which might explain the re-appearance of L1ORF1p in round spermatid nucleus and then in elongating spermatid cytoplasm, but not in P/D. (in this study, and [130]).

An intriguing and still open question is that of a purpose or function of TRDM in meiosis. Some existing data allow us to consider several possibilities. First, is that TRDM is merely a byproduct of remodeling of meiotic chromosomes in preparation to subsequent events in MPI. Protracted pre-meiotic S-phase is not only the time of the last round of DNA replication before meiosis but also is the time of loading of numerous meiosis-specific cohesin complexes onto chromosomes [191, 192]. Conceivably, these and other changes might necessitate uncoupling of maintenance DNA methylation from DNA replication to allow efficient meiotic chromosome remodeling in preparation for subsequent meiotic events. In this regard, TRDM fulfills a role of a sensitive marker informing us of as yet not well-understood changes to meiotic chromosomes.

Second, could reducing DNA methylation facilitate key MPI processes including homolog search and synapsis, meiotic chromosome structure and meiotic recombination? This idea is supported by prior studies in fungi and plants demonstrating strong effects of DNA methylation on the genomic landscape of meiotic recombination [193-195]. DNA hypomethylation was also found to correlate with increased recombination rates in Prader-Willi/Angelman syndrome imprinted region [196]. Of note, is that in the female germline, gonadal PGCs enter MPI at the point of their lowest levels of genome-wide DNA methylation. Thus, the identification of TRDM at the onset of meiosis in the male germline suggests that germ cells of both sexes and not just fetal oocytes enter meiotic prophase with lowered levels of DNA

methylation. This similarity strengthens the notion that reducing DNA methylation levels might be required for normal progression of meiotic processes.

Comparison of DNA methylation dynamics and transposon expression between wild type and *Mael*$^{-/-}$ mice points to the critical role of preexisting DNA methylation levels for the successful progression and completion of meiosis. Thus, another possibility is that TRDM contributes to germ-cell quality control by exposing genomes with insufficient DNA methylation of transposons. This idea builds upon observations of transient L1ORF1p expression in wild-type mice and massive L1ORF1p overexpression in piRNA or DNA methylation-deficient mutant animals. The combination of DNA methylation and RNA expression analyses suggests that TRDM brings germ cells close to widespread reactivation of transposons thus necessitating extensive DNA methylation in pre-meiotic germ cells. It is also important to note, that this potential role of TRDM is reminiscent of events during fetal oogenesis where differential L1 expression drives selective elimination of oocytes with excessive L1 levels [197]. Importantly, regardless of the specific functional role of TRDM in meiosis, this work reveals an additional step in epigenetic programming of the male germ cells and contributes to our understanding of the epigenetic context of meiosis.

Acknowledgments

We want to thank: Fred Tan for continuous help with troubleshooting during bioinformatics analysis; Allison Pinder for sequencing and maintaining a high-standard sequencing core facility; Eugenia Dikovsky and staff for taking care of mice and maintaining excellent mouse facility; Svetlana Deryusheva, for critical reading and constructive feedback on the manuscript submitted for publication based on the data from this chapter; Safia Malki, for thoughtful comments and proofreading of the submitted manuscript; Marla Tharp for thoughtful corrections of the final submitted manuscript.

Final Short Summary

My work shows that when postnatal male germ cells enter meiosis, they undergo genome-wide transient relaxation of DNA methylation (TRDM). We find that in the course of TRDM, germ cells briefly release transposable elements from repression. Our data suggest that genome-wide reduction of DNA methylation levels during TRDM occurs by a passive, DNA replication-coupled mechanism. In contrast to robust maintenance DNA methylation within minutes after DNA synthesis in mitotically dividing cells, remethylation of meiotic chromosomes is a gradual and uneven process across the genomic landscape. TRDM may create a permissive environment for the efficient reciprocal exchange of genetic information between parental chromosomes in meiosis or ensure gamete quality. My work has an important implication for understanding why many mutations, like the one in Maelstrom, exhibit sensitivity and subsequent germ cell demise in MPI. Namely, my research implicates the epigenetically relaxed state of meiotic onset in this sensitivity.

Chapter 4
Materials and Methods

4.1 Mice

Adult C57BL/6J male mice (2–5 month old mice) (Jackson Laboratory) were used as a source of adult testes. All experimental procedures were performed in compliance with ethical regulations and approved by the IACUC of Carnegie Institution for Science. The *Mael* knock-out strain was described in [116].

4.2 Germ cell isolation

Adult testes were dissected out of mice and processed for Fluorescence Activated Cell Sorting (FACS) as described in Chapter 2 of this thesis for WT and *Mael-/-* mice [28, 146]. FACS-enriched WT germ cells included Spg and PL, L, Z, P and D spermatocytes. Only the Spg, PL and a combined LZp cell populations were FACS-enriched from *Mael-/-* testes. Testicular somatic cells were obtained from both, WT and *Mael-/-* mice [28, 146]. Cell purity after sorting was evaluated based on immunofluorescence analysis of nuclear spreads with antibodies against proteins enriched in either pre-meiotic Spg (DMRT1, DMRT6) [119, 120] or meiotic cells (SYCP3, yH2AX), as described before [28, 146]. We have determined that our Spg population is enriched in differentiated, rather than undifferentiated Spg. This conclusion is based on the absence or low abundance of DMRT1, presence or high abundance of DMRT6, and presence of heterochromatic foci, as evaluated by DAPI staining. Note, that the collected PL cells were negative for DMRT1, however, some cells contained DMRT6 puncta. Cell counts based on the detailed IF analysis, as described before [28, 146], were conducted with sorted cell fractions used for subsequent analysis exhibiting a purity of >85% for Spg, 75-85% for PL, 65-85% for L, >80% for Z, > 90% for P, >90% for D.

4.3 Immunofluorescence (IF)

IF on testicular sections or meiotic spreads was performed as described before [28, 146].

Briefly, a 50 µl aliquot of sorted cells was mixed with 50 µl hypotonic buffer (30 mM Tris-Cl, 50 mM sucrose, 17 mM sodium citrate dehydrate, 5 mM EDTA in water) and incubated at RT for 30 min. Cells were pelleted at 5000 rpm for four minutes, 80 µl supernatant was removed and cells re-suspended with 65 µl of 100 mM sucrose, pH 8.2. Next, 30 µl of the suspension was applied to a glass slide (Superfrost Plus, VWR) dipped just before in a solution of 1% paraformaldehyde, pH9.2, supplemented with 0.1% Triton X-100. The pH of all solutions was set using 50 mM sodium borate. Nuclei were dried overnight (O/N) in a slightly opened humid box over water. Finally, the slides were washed for one minute in 0.2% Photoflo (Kodak), dried at RT, and stored at minus 20°C. Frozen slides were thawed at 42°C, followed by serial washes, five minutes each, of 0.5% Triton X-100 in 1X PBS, 0.05%Triton X-100 in 1X PBS and 1X PBS. Slides were treated with blocking buffer (2% BSA, 0.05% Triton X-100) for 30 minutes and incubated O/N at RT with primary antibodies (1:750 anti-SCP3, Abcam, #ab15093 and 1:1000 anti-phospho-Histone γH2A.X, Millipore, #05-636) in blocking buffer supplemented with 10% normal donkey serum). After washes as above, cells were incubated with secondary antibodies (1:500 donkey anti-rabbit 594, Life Technologies, #A21207 and 1:1000 donkey anti-mouse 488, Life Technologies, #A21202). Slides were subsequently washed as described above and counterstained for 5 min in 0.1 µg/ml 4,6-diamidino-2-phenylindole (DAPI) in PBS. After a rinse in PBS, the slides were mounted in Vectashield (Vector Laboratories) and sealed with nail polish. To quantify percent purity, between 100 and 250 spread cells from each sorted cell population, were scored and classified after each sort.

A laser-scanning confocal microscope (Leica DM6000, Exton, PA or Zeiss) was used for imaging cryosections. For imaging meiotic spreads, an upright fluorescence microsope was used (Olympus BX61). All images were analyzed in ImageJ.

The following primary antibodies were for immunofluorescence (IF): monoclonal anti-γH2AX (Mouse, 1mg/ml, 1:1000, Millipore 05-636), polyclonal anti-SYCP3 (Rabbit, 1mg/ml, Abcam, ab15092. IF: 1:500 dilution), polyclonal anti-ORF1p (Rabbit, 1mg/ml, a kind gift from

Dr. Martin. IF: 1:500 dilution) [130], monoclonal anti-DMRT1 (Mouse, 200 μg/ml, Santa-Cruz, sc-10222. IF: 1:200 dilution), polyclonal anti-DMRT6 (Rabbit, a kind gift from Dr. Zarkower. IF: 1:200 dilution) [120], monoclonal anti-sp56 (Mouse, Pierce, MA1-10866. IF: 1:750), anti-DNMT3a, anti-DNMT3A2, and anti-DNMT3B (rabbit, a kind gift from Shoji Tajima. IF: 1:1000-1:5000 dilution). The following secondary antibodies (2mg/ml) were used in this study: donkey anti-rabbit Alexa Fluor 594 (1:500), donkey anti-rabbit Alexa Fluor 488 (1:500 for SCP3, ORF1p and 1:200 for DMRT6), donkey anti-mouse Alexa Fluor 488 (1:1000 for yH2AX, 1:200 for DMRT1).

4.4 Cryosections

After dissection of the testis, the tunica was removed, the testis was fixed (2% PFA in PBS) at 4C for 4 hours, shaking. Samples were passed through a sucrose (dissolved in water) gradient (10% for 1h, 20% for 1h, 30% overnight at 4C), embedded in OCT and stored at -80 C. Sections of 10 μm were used for IF.

4.5 EdU labeling

Adult mice 1-3 months old were injected with 12.5μg/g of body weight EdU (0.5mg/ml DMSO stock) dissolved in 200μL water. Mice were sacrificed 2 hours after injection and processed for FACS or for cryosections as described above. EdU detection with Click-iT EdU Alexa Fluor Kit was performed as described in the manual (Invitrogen).

4.6 Whole Genome Bisulfite Sequencing (WGBS)

WGBS Background. WGBS is a process whereby input DNA is treated with sodium bisulfite and sequenced. The sodium bisulfite method, development by Hayatsu [198], involves the conversion of cytosine (C) to uracil (U), while methylated C (5mC) is protected because it does not react with bisulfite. Mouse chromosomes contain 42% mean (G + C) content and less than

5% 5'-CpG-3' (CpG) content [68]. The Cs that occur in CpG dinucleotide context are often highly methylated in mammals, while non-CpG cytosines are predominantly unmethylated. Thus, since most Cs in the genome are not methylated, after bisulfite conversion of Cs into Us, followed by the amplification of bisulfite-treated DNA by PCR and sequencing, unmethylated Cs are replaced with Ts resulting in a nearly C-less genomic sequence [148]. By comparing the modified DNA with the original reference sequence, the methylation state of the original DNA can be inferred. Thus, bisulfite conversion combined with next-generation sequencing can accurately measure DNA methylation at specific genomic loci, at single-base resolution. Subsequently, computational tools are used to map BS-seq reads to the genome, determine the methylation state of any given C and detect and measure differential methylation between two samples. In this study, we use *Bismark* software for BS-Seq data alignment and methylation calling [147, 148] and *bsseq* software package [156] for accurate estimation of DNAme at single CpG level and statistically principled differential methylation analysis. An extended commentary and computational protocols on every aspect of BS-Seq analysis can be found in the Methods and Procedures section of this chapter.

WGBS experimental setup. Each biological replicate consisted of pooled cells from 2-3 different animals from different FACS procedures. Two biological replicates (2x) were used for WT samples Spg, PL, L, Z, P, D and epididymal spermatozoa (Spz). Spz was obtained by pushing sperm out of the epididymus. For *Mael*-/-, 1x Spg, 3x PL and 2x LZp biological replicates were used.

Genomic DNA (gDNA) preparation. gDNA was prepared as follows. FACS-enriched cells stored at -80 C in nuclease-free 1.5 ml Eppendorf tube, were hand-homogenized in 100µl tail lysis buffer using disposable pellet pestle (Kimble Kontes). Up to 300µl tail lysis buffer, and 5µl of Proteinase K (Life Technologies, 20mg/ml) were added to the samples and the samples were

incubated at 55 C for 2-3 hours. At the end of lysis, 2µl of Linear Acrylamide (Ambion, 5mg/ml) were added to samples. DNA was extracted with Phenol Chloroform Isoamyl Alcohol (25:25:1, pH 8.05-8.35) (Life Technologies, #15593) using Phase Lock Gel (PLG) Light tubes (5 Prime). One microliter of RNase A (Thermo Scientific, #EN0531, 10 mg/ml) was added to the aqueous phase, and the samples were incubated at 37 C for 30 min, transferred to a PLG tube and mixed with chloroform. The aqueous phase was processed for DNA precipitation using Ethanol. Prior to the addition of sodium acetate, 2 µl more of Linear Acrylamide was added to the sample. DNA pellet was resuspended in 15-20µl EB (pre-heated at 65 C) and kept overnight at room temperature. The DNA was quantified with picoGreen assay (Molecular Probes) using SpectroMax microplate spectrophotometer.

WGBS library preparation. The library was prepared based on Illumina's 'WGBS for Methylation Analysis' protocol. Isolated mouse gDNA was spiked with approximately 0.1% unmethylated cl857 Sam7 Lambda DNA (Promega) and sheared to fragments with a range of 200-600 bp using a Covaris M220 Ultrasonicator. The diluted fragmented gDNA was concentrated and end-repaired. Afterwards, the samples were processed for adenylation and adaptor ligation. Methylated Illumina adapters with indexes AD001-AD016, AD018-AD023, AD025 and AD027 were used. The adaptor-ligated DNA fragments were processed for bisulfite conversion according to the manufacturer's protocol (EZ DNA methylation Gold kit, Zymo Research). Bisulfite-treated DNA underwent 15 rounds of PCR amplification. The PCR amplified samples were used for Illumina cluster generation and sequencing using Illumina *TruSeq* RNA Sample Prep Kit v2 components. The BSC libraries were sequenced on Illumina HiSeq2000 platform, yielding 100bp Paired-End (PE) reads. WT biological replicate one was sequenced at a different time from WT biological replicate number two. All *Mael-/-* samples, Spz and somatic cells were sequenced at once. Sequencing was performed by Allison Pinder in our core facility. Each sample was run in a single lane, spiked with 5% Illumina PhiX genomic DNA control to

increase library diversity, as per *Illumina* recommendation. Data were downloaded onto our servers in FASTQ format for processing.

WGBS read alignment and extraction of DNA methylation evidence. We used *Bismark* program [147, 148] for mapping bisulfite-converted reads to mouse genome assembly NCBI37/mm9 (**Table 3-1**). For estimation of bisulfite rate conversion, reads were aligned to Lambda genome. Bismark version 0.13.0 was used, except where noted otherwise. The alignment was performed with respect to the bisulfite-treated Watson (original top) and Crick (original bottom) strands, and not their reverse complements, as the library was prepared in a strand-specific (directional) manner [147]. No trimming was performed prior to alignment [199]. After alignment *Bismark* de-duplication module was used to remove PCR duplicates, defined as reads that have the same orientation, and start and end at the same position. Of the duplicated reads, the remaining read is randomly chosen (**Table 3-1**).

After alignment and de-duplication of reads, we used *Bismark* to extract and summarize CpG methylation evidence present in the unique alignments. A piece of CpG "evidence" is an alignment that overlaps cytosine position of a CpG in the reference genome and was either a T (indicating a lack of methylation) or a C (indicating presence of methylation) in the alignment. After extraction of CpG (and non-CpG) evidence, the CpG evidence was filtered based on evaluation of methylation bias (M-bias) plots. Based on the M-bias plot, we excluded the first 6 nt from 5' end of read 1 and 10 nt from read 2, and the last 1 nt from 3' end of both reads prior to the extraction of methylation. Subsequently, we extracted CpG coverage into a file containing information for both strands. Finally, we merged strand-specific information. The final output text file contained chromosome (chr) name, chr start, chr end, CpG methylation percentage, count C and count unmethylated C. The results are summarized in **Table 3-2**, while **Protocol 1** summarizes computational steps taken to perform this analysis. **Appendix contains scripts 1-7** used to process fastq files into final file containing CpG methylation evidence. These final DNA

methylation files were then typically analyzed with *bsseq* package. The files were imported into *bsseq* using *read.bismark()* function. For the analysis of all mappable CpGs per sample, mean coverage and "raw" DNA methylation levels were calculated using *getCoverage()* and *getMeth()* functions, respectively in bsseq.

Bioinformatics analysis of global DNA methylation levels. Correlation between replicates of WGBS data **(Table 3-3)** was performed as follows. Biological replicates were compared pairwise (e.g., Spg1 with Spg2). Final *Bismark* output files containing CpG methylation and coverage were imported into R using <read.bismark> function from *bsseq* package. Once imported, these files are considered *bsseq* format files. DNA methylation was extracted and summarized in non-overlapping bins of 500 CpGs using *rep()* function, followed by aggregation of data and computation of mean values, using *aggregate()* function in R. Pearson correlation coefficient was then calculated using *cor()* function in R.

Global DNA methylation analysis was performed using *bsseq* package [156]. For the analysis of WT samples, two replicate groups were formed. Each one consisted of 7 samples (Spg, PL, L, Z, P, D and Spz) and made up a single *Bsseq* object. Only those CpGs that were covered by at least one read in all samples (common CpGs) were analyzed. The summary of DNA methylation analysis for CpGs in common, between samples of an individual biological replicate group, can be found in **Table 3-4** and **Protocol 2** outlines the steps taken to perform this analysis in *bsseq*. Since only these common CpGs were involved in data analysis, the overall DNA methylation levels for each sample slightly differs from "raw" DNA methylation values obtained for all mappable CpGs of a given sample.

For plotting DNA methylation across the length of the chromosome (e.g., **Fig 3-3B**), a custom python script was used and DNA methylation was averaged using sliding non-overlapping windows of 100 kbp.

Annotation used for DNA methylation analysis. For most annotation, we used UCSC genome browser database (http://genome.ucsc.edu). The following genomic feature annotations were directly extracted from UCSC Table Browser based on mm9 genome: introns, exons, CpG Islands and repeats annotated by RepeatMasker (RMSK) (http://repeatmasker.org). Promoter coordinates were extracted from the UCSC/knownGene transcriptome file by taking +1kb to -1kb relative to the TSS, in a strand-conscious manner. Intergenic regions were extracted from the UCSC Table Browser by (1) extracting BED file containing 'chrom, txStart and txEnd' via 'selected fields from primary and related tables' output format (2) uploading the new file as a custom track and specifying 'custom track' in the <group> category of table browser (3) using <intersection> to create and intersection of the custom track with itself, selecting the 'Base-pair-wise intersection (AND) of User Track and User Track' option for intersection, and 'Complement User Track*'. The geneID and geneSymbols were obtained from specifying the selected fields in the output format and selecting knownGene (name), and kgXref (geneSymbol) fields. A custom script was used to change ucsc_ids to geneSymbols in the BED files.

Analysis of sequencing coverage after WGBS. For chromosome-based visualization, deduplicated bam files from BS-seq were sorted using Samtools. To make the simultaneous analysis of multiple files for each MPI stage, Spg and Spz possible, individual chromosomes were analyzed separately (extracted from sorted bam files). Biological replicates were analyzed separately for independent assessment. The files were uploaded into SeqMonk and read coverage was quantitated in the following manner: running, non-overlapping window probes of 5kb were created to span the chromosome length. Read counts (the probes) were quantitated using the SeqMonk's *Read Count Quantitation* approach where we counted all reads and corrected for total read count based on the largest data set. For overall coverage quantitation, running, non-overlapping window probes of 5kb were created to span the chromosome length. Probe read counts were quantitated using the SeqMonk's *Read Count Quantitation* approach where we

counted all reads and corrected for total read count based on the largest data set. Data store summary report was exported.

Determination of overlap between datasets. Generally, overlaps were computed using *bedtools intersect*. For example, for intersecting DNA methylation with late replicating regions, <intersect –wa –wb> was used, with replication timing (RT) file (-a) and DNA methylation file (-b). The RT file contained <chr RT/start RT/end RT/RT value>. The CpG methylation file contained <chr C/start C/end C/methylation level. Pearson correlation between DNA methylation values and RT values was performed using *cor()* function in R.

Analysis and Annotation of Differentially Methylated Regions (DMRs). DMR analysis was performed using *Bsseq* package with previously optimized settings for DMR blocks [156, 199]. *Bsseq* employs local likelihood method, aggregating information from neighboring CpGs in a coverage-conscious manner and uses the combined data from two biological replicates to estimate DNA methylation at single CpG level. For this analysis we required that each CpG be covered at least once in all 4 samples compared pairwise (two biological replicates per two stages). This selection resulted in a median CpG coverage of 3X-7X per sample (or 6X-14X per duplicate) and an overall coverage of > 77% of all genomic CpGs (**Table 3-7**). **Protocol 3** contains pipeline for the computational analysis of DMRs used in this study.

For analysis of DMRs and replication timing, genomic coordinates of hypomethylated regions were obtained from DMR blocks and intersected with early or late replication timing (RT) coordinates. For every region within A (early or late RT domain) a number of intersections with B (DMR block) were computed. To calculate the proportion of an overlap between DMR blocks formed between Spg and PL ('WTSpgPL' DMRs) and all other DMR blocks, <*bedtools intersect* –wa –wb> was utilized, using a file containing WTSpgPL DMRs as (-a) and separate files, each containing DMRs between WT PL and L, L and Z, Z and P, P and D as (-b).

Subsequent processing of the output involved adding the number of all intersections that matched WTSpgPL DMRs and normalizing to the total number of DMRs for a particular pairwise comparison.

Annotation from Illumina's iGenomes (genes.gtf), based on the RefSeq dataset (dating July, 2015), was used for annotating DMRs with genes.

To measure enrichment of overlap between transcriptionally upregulated genes and hypomethylated promoter DMRs, we measured a ratio of overlap between them and compared the ratio to the ratios obtained for upregulated genes outside promoter DMRs and genome-wide (**Table 3-9**). If the overlap ratio for genes inside promoter DMRs was same or similar to the other ratios then, an overlap was considered random. Subsequently, we used Fisher's exact test to examine strength of overlap between DMRs and gene transcription. For each set of DMRs, a custom python script was used to form a 2x2 table containing significantly upregulated genes that overlapped with DMRs, significantly upregulated genes that fell outside of DMRs, all genes found inside of the DMRs, and all genes found outside of the DMRs. To evaluate the significance of overlap, we calculated p-values using Fisher's exact test.

4.7 RNA-Sequencing

Total RNA preparation. Total RNA was isolated from FACS-enriched fractions from adult C57BL6 male mouse testis. In most cases, due to the limited availability of enriched cells, total RNA from 2-4 mice (2-4 independent FACS enrichment sessions) was pooled to create one sample. Moreover, early MPI cells including PL, L and Z are not very transcriptionally active, compared to the cells of late MPI, the P and D. For example, P spermatocytes, have been recently estimated to produce relatively large amounts of RNA per cell (~12 pg) similar to an average liver cell (~15pg) [172]. Thus, while it is straightforward to collect biological replicates for late MPI, similar amounts of RNA for early MPI are obtained only from multiple independent FACS sorts. Thus, for WT, six samples were obtained, and included Spg, PL, L, Z and P. For *Mael-/-*, 3

samples were obtained, and included Spg, PL and LZp. Mael KO LZp was collected as one fraction since the defects observed in Mael KO MPI cells make it difficult to differentiate L and Z and Z-like and P-like cells [116, 146]. 1ul of RNA was used for evaluation on the BioAnalyzer. For *RNA-seq*, one biological replicate was used for WT Spg, PL, L, Z, P, D, Soma and Spz; For *Mael-/-* 2x Spg, 3x PL, 1x LZp and 2x Soma were used.

rRNA depletion of total RNA. Ribosomal RNA (rRNA) was removed from total RNA (up to 50ng) using Ribo-Zero Gold rRNA Removal Kit according to the manufacturer's protocol.

cDNA library preparation. The TruSeq RNA Sample Preparation Kit v2 was used to prepare cDNA library from ribosomal RNA depleted RNA. The libraries were prepared by Allison Pinder of Carnegie's Sequencing Core facility as described in the manufacturer's protocol (Pub. Part no.: 15026495) following low sample protocol. DNA fragments were enriched with PCR for 15 cycles. One microliter of the resulting library was used for validation and quantification analysis, using Agilent Technologies 2100 Bioanalyzer and Agilent DNA-1000 chip. The cDNA libraries were sequenced as single end 50-mers using the Illumina HiSeq2000 platform, yielding a total of ~ 246 million reads (26-66 million total reads per sample).

RNA-seq quality validation and read mapping. The quality of the raw RNA-seq libraries was evaluated using fastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The fastQC-reported "Per base sequence quality" measure was very good, with more than 92% of all reads having a quality score of more than 30, and mean quality score of more than 36. This score is based on Phred-scale quality, where a score of 20 corresponds to one error in every 100 base calls, or 99% accuracy, and all of our samples exhibit a very high score across all 50bp, with only minor decrease in quality at position 50. Although the " per base sequence content" exhibited a characteristic to Illumina RNA-Seq bias in the first 11 bases (related to library preparation steps),

we did not trim away these nucleotides, since their overall quality was still high, and the trimming method has been shown to be most beneficial for intermediate quality threshold of 20 to 30, whereas all, except the last basepair in our data exhibit per base sequence quality of 30 or above.

RNA-seq read alignment. The read alignment was performed with TopHat (v2.0.7) [160, 200] (https://ccb.jhu.edu/software/tophat), using short read mapping program Bowtie 2 (v2.0.6). During the alignment, we provided a transcriptome file that contained gene annotation. The reads were processed based on NCBI37/mm9 mouse genome and UCSC RefSeq gene annotation obtained from Illumina iGenomes (dating July, 2015). Alignment statistics are found in **Supplementary Table**.

Bioinformatics analysis of RNA-seq. We used HTSeq package to count sequencing reads that overlap with gene transcriptome (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html) [159]. Specifically, we used <htseq-count –s no –a 10 input.sam iGenomes.gtf> command. The output is a tab-delimited text file containing counts for each gene (gene id and number of read counts). Subsequently, to evaluate differential expression we used edgeR [201]. Specifically, we (1) built a counts table with all samples, using *DGElist* function, (2) normalized counts using the default TMM method, which accounts for compositional differences between the libraries, using *calcNormFactors* function (3) obtained a table with normalized count-per-million (CPM), using *cpm* function, which we used directly for data analysis, or converted CPM to RPKM by (cpm/gene length/1000). Gene length file was kindly provided to us by Xiaobin Zheng. For the differential expression analysis, an exact test was performed with an estimated Biological Coefficient of Variation (BCV) of 0.1, and *topTags* function was applied. A final table containing logFC (is log2FC), logCPM (is log2CPM), p-value and FDR value for each gene was obtained. See **Appendix (scripts 8-10)** for scripts used to process fastq files into the final file containing RNA read counts for transcriptome.

Gene Annotation and Differential Expression analysis. One widely used data annotation is NCBI's RefSeq Refgene dataset (Pruitt et al, 2005). Illumina provides ready-to-use reference sequence annotation as part of its iGenomes collection (https://support.illumina.com/sequencing/sequencing_software/igenome.html) that is based on the RefSeq dataset, but is more conveniently organized. For example, the iGenomes annotation (GTF) file already contains gene names and other attributes such as transcription start site id (tss_id) and transcript_id, avoiding the need to cross-reference in order to get this information otherwise. Thus both, the alignment and differential expression (DE) analysis were performed using the most updated version of NCBI's gene annotation.

Analysis of transposable element RNA. The file containing mouse genome repetitive elements was obtained from RepeatMasker and downloaded as mm9.fa.out.gz file (http://www.repeatmasker.org). For the analysis of the transcriptional landscape of repetitive elements we used RepEnrich [184] according to the suggested protocol (https://github.com/nerettilab/RepEnrich). Briefly, we aligned RNA-seq data to the genome using bowtie1 parameters that allow only unique mapping (-m1) and outputted multimapping and uniquely mapping reads into separate files. We ran RepEnrich python script on the data and then used EdgeR for subsequent processing of fraction counts file, which contained 1444 repetitive element entries. Specifically, we (1) built a counts table with all samples, using *DGElist* function, (2) normalized counts using the default TMM method, which accounts for compositional differences between the libraries, using *calcNormFactors* function (3) obtained a table with normalized count-per-million (CPM), using *cpm* function, which we used directly for data analysis. For the differential expression analysis, an exact test was performed with an estimated dispersion specific to each pairwise comparison (which varied between 0.06 and xxx), and *topTags* function was applied. A final table containing logFC (is log2FC), logCPM (is log2CPM), p-value and FDR value for each repetitive element entry (subfamily).

Protocol 1. WGBS read alignment and extraction of DNA methylation evidence.

This document contains a pipeline for alignment of the 100-by-100 bp HiSeq 2000 paired end bisulfite sequencing reads using Bismark bisulfite mapper and methylation caller, and Bowtie2. Mouse mm9 build and the genome for lambda phage were used. The protocol uses only one sample, “SampleVG2”, for demonstration. Supplementary Tables 1-3 summarize all the results.

Quick References:
FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
Bismark (http://www.bioinformatics.babraham.ac.uk/projects/bismark/)
Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/)

Bismark version v0.13.0 was used
Bismark_methylation_extractor experimental version (11 December 2014) by Felix Kruger was used, obtained via personal communication prior to its official release in v0.13.1

---

1. Perform Quality Control on raw sequence data using fastqc on a subset of reads

<fastqc> assesses an overall quality of the bisulfite-converted reads. Particularly informative are “Per Base Sequence Quality” and “Per Base Sequence Content” plots.

```
fastqc VG2_NoIndex_L001_R1_001.fastq.gz VG2_NoIndex_L001_R2_001.fastq.gz
```

* R1 refers to Read 1, R2 refers to Read 2

2. Concatenate all reads of the same sample into one file, each paired end independently

```
cat /Sample_VG2/VG2_NoIndex_L001_R1_0??.fastq.gz > VG2_IDX2WTSPG2_L001_R1ALL.fastq.gz
cat /Sample_VG2/VG2_NoIndex_L001_R2_0??.fastq.gz > VG2_IDX2WTSPG2_L001_R2ALL.fastq.gz
```

3. Prepare genome of interest for bisulfite alignments

Bismark will create two folders within the specified directory, one for a C --> T converted genome, and another for the G--> A converted genome. Only one genome (genome.fa) should be in a folder specified.

USAGE: bismark_genome_preparation [options] <path_to_genome_folder>

```
#Prepare mm9 genome
/mnt/sequence/gaysinskaya/bin/bismark_genome_preparation --bowtie2
/mnt/sequence/gaysinskaya/Bismark_bt2mm9

#Prepare lambda phage genome
/mnt/sequence/gaysinskaya/bin/bismark_genome_preparation --bowtie2
/mnt/sequence/gaysinskaya/Bismark_bt2lambda
```

4. Align bisulfite genome using Bowtie2

USAGE: bismark [options] <genome_folder> -1 <mates1> -2 <mates2>

```
# --bowtie2, Uses Bowtie 2
# --bam, The output will be written out in BAM format
# -p, Launch parallel search threads (default: 1)
# --score_min, minimum alignment score needed for an alignment to be considered
"valid"
```

```
# -1 <mates1>, Specify paired-end reads 1
# -2 <mates1>, Specify paired-end reads 2
# -o, indicates output folder name, otherwise, current folder

Can first test the alignment using a subset of 10000 using  $bismark -u 10000

#Align to mouse mm9 genome
bismark --bowtie2 --bam -p 4 --score_min L,0,-0.4
/mnt/sequence/gaysinskaya/Bismark_bt2mm9 -1 VG2_IDX2WTSPG2_L001_R1ALL_val_1.fq.gz -2
VG2_IDX2WTSPG2_L001_R2ALL_val_2.fq.gz


#Align to lambda phage genome
bismark -q --bowtie2 --bam -p 4 --score_min L,0,-0.4 -o lambda
/mnt/sequence/gaysinskaya/Bismark_bt2lambda -1 VG2_IDX2WTSPG2_L001_R1ALL_val_1.fq.gz -
2 VG2_IDX2WTSPG2_L001_R2ALL_val_2.fq.gz
```

5. Deduplicate reads

<deduplicate_bismark> removes alignments to the same position in the genome which can arise by e.g. PCR amplification. Paired-end alignments are considered a duplicate if both partner readd start and end at the exact same position. Deduplication allows only 1 read for each position in the genome.

```
deduplicate_bismark --paired --bam
VG2_IDX2WTSPG2_L001_R1ALL_val_1.fq.gz_bismark_bt2_pe.bam
```

6. Examine M-bias table and plot to determine methylation bias

Methylation state should not depend on read position. Methylation-bias (M-bias) plots shows methylation levels across the read positions and ought to show an overall flat (albeit with noise-like variation) horizontal line. See Kasper D Hansen et al, Genome Biology 2012 for detail.

USAGE: bismark_methylation_extractor [options] <filenames>

```
bismark_methylation_extractor --multicore 4 --mbias_only --paired-end
VG2_IDX2WTSPG2_L001_R1ALL.fastq.gz_bismark_bt2_pe.deduplicated.bam
```

7. Extract methylation, while eliminating M-bias

Based on M-bias analysis, we removed 6 nucleotides from the 5' end of read 1, 10 nucleotides from the 5' end of read 2, and 1 nucleotide from 3' end of both reads

USAGE: bismark_methylation_extractor [options] <filenames>

```
# --multicore <int>, Sets the number of cores to be used for the process.
# --gzip, Methylation files will be written out in a compressed GZIP form
# --paired-end, Indicates that input files are paired-end
# --no_overlap, For paired-end reads it is theoretically possible that Read 1 and Read
2 overlap. This option avoids scoring overlapping methylation calls.
# --report, Prints out a methylation summary and the parameters used
# --ignore <int>, Ignore the first <int> bp from the 5’ end of Read 1
# --ignore_r2 <int>, Ignore the first <int> bp from the 5' end of Read 2
# --ignore_3prime <int>, Ignore the last <int> bp from the 3' end of Read 1
# --ignore_3prime_r2 <int>, Ignore last <int> bp from the 3' end of Read 2
# -o, Specifies name of the existing output folder
# --comprehensive, Merge all strand-specific methylation info into 3 context-dependent
output files: CpG-context, CHG-context and CHH-context
```

```
bismark_methylation_extractor --multicore 2 --gzip --paired-end --no_overlap --report
--ignore 6
--ignore_r2 10 --ignore_3prime 1 --ignore_3prime_r2 1 -o methextract --comprehensive
/mnt/sequence/gaysinskaya/FACS_BSSEQ2014_Rep2/Sample_VG2/rawalign
VG2_IDX2WTSPG2_L001_R1ALL.fastq.gz_bismark_bt2_pe.deduplicated.bam

#View output of interest: Example for cytosines in CpG context

gunzip
CpG_context_VG2_IDX2WTSPG2_L001_R1ALL.fastq.gz_bismark_bt2_pe.deduplicated.txt.gz
head CpG_context_VG2_IDX2WTSPG2_L001_R1ALL.fastq.gz_bismark_bt2_pe.deduplicated.txt
```

The bismark_methylation_extractor output is tab-delimited in the following format (1-based coords):

```
<seq-ID> <methylation state> <chromosome> <start position (= end position)>
<methylation call>

Bismark methylation extractor version v0.14.3_devel
HWI-ST375:284:C5T2AACXX:1:1101:1132:1990_1:N:0:      +      chr8   117459180 Z
HWI-ST375:284:C5T2AACXX:1:1101:1132:1990_1:N:0:      +      chr8   117459136 Z
HWI-ST375:284:C5T2AACXX:1:1101:1132:1990_1:N:0:      +      chr8   117459111 Z
HWI-ST375:284:C5T2AACXX:1:1101:1203:1989_1:N:0:      +      chr6   111824998 Z
HWI-ST375:284:C5T2AACXX:1:1101:1203:1989_1:N:0:      +      chr6   111824917 Z

* Methylated cytosines receive a '+' orientation, * Unmethylated cytosines receive a
'-' orientation
* z     unmethylated C in CpG context
* Z     methylated C in CpG context
```

8. Generate strand-specific coverage file

```
<bismark2bedGraph> module will write out a coverage (.cov) file, and an optional
bedGraph (.bedGraph) output. By default, only cytosines in CpG context are considered.

USAGE: bismark2bedGraph [options] -o <output> [methylation extractor input files]

bismark2bedGraph -o VG2_IDX2WTSPG2
CpG_context_VG2_IDX2WTSPG2_L001_R1ALL.fastq.gz_bismark_bt2_pe.deduplicated.txt.gz

head VG2_IDX2WTSPG2.bismark.cov
```

The CpG coverage report is tab-delimited (1-based coords):

```
<chromosome> <start position> <end position> <methylation percentage> <count
methylated> <count unmethylated>

chr8         3000264             3000264      100    1      0
chr8         3001031             3001031      100    1      0
chr8         3001165             3001165      0      0      1
chr8         3001720             3001720      100    1      0

* strand specific information per CpG is in separate line
```

9. Generate final strand-merged coverage file

```
<coverage2cytosine> module generates a cytosine methylation report for a genome of
interest and a sorted methylation input file (.cov) produced by the script
"bismark2bedGraph". By default, the output uses 1-based chromosome coordinates and
reports CpG positions only (for both strands individually and not merged in any way).

The option --merge_CpG will generate and post-process the genome-wide report to write
out an additional coverage file which has the top and bottom strand methylation
evidence pooled into a single CpG dinucleotide entity. This is the desirable input
format for downstream processing with R-package bsseq (by K.D. Hansen).
```

```
 USAGE: coverage2cytosine [options] --genome_folder <path> -o <output> [input]

coverage2cytosine -o VG2d --merge_CpG --genome_folder
/mnt/sequence/gaysinskaya/Bismark_bt2mm9 VG2_IDX2WTSPG2.bismark.cov

head VG2d
```

The genome-wide cytosine report is tab-delimited (1-based coords):
<chromosome> <position> <strand> <count methylated> <count unmethylated> <C-context> <trinucleotide context>

```
chr8         3000264            +      1      0      CG     CGT
chr8         3000265            -      0      0      CG     CGT
chr8         3000685            +      0      0      CG     CGT
chr8         3000686            -      0      0      CG     CGA
```

The genome-wide cytosine report is tab-delimited in the following format (1-based coords):
<chromosome> <start position> <end position> <methylation percentage> <count methylated> <count unmethylated>

```
head VG2d.merged_CpG_evidence.cov
chr8         3000264            3000265      100.000000    1      0
chr8         3001031            3001032      100.000000    1      0
chr8         3001164            3001165      0.000000      0      1
chr8         3001719            3001720      100.000000    1      0
```

* strand specific information per CpG is merged into one line
*** The merged_CpG_evidence.cv is the final file used for all subsequent analysis of CpG methylation**

Protocol 2: Analyzing WGBS with the bsseq package

This document contains a pipeline for analyzing WGBS data with Bsseq package, pre-processed by Bismark. The document focuses on analysis-related tasks and questions. The document does not deal with differential methylation, but uses bsseq package to perform basic, overall DNA methylation analysis. Supplementary Table 5 and Figure 2A summarize the results.

Quick References:
Bsseq (http://bioconductor.org/packages/release/bioc/html/bsseq)

R version 3.0.1 (2013-05-16) was used
Bsseq version 0.10.0 was used

Sample IDs:
Group 1: VG1d (Spg1), VG3d (PL1), VG5d (L1), VG7d (Z1), VG911d (P1), VG12d (D1)m VG24d (Spz1)
Group 2: VG2d (Spg2), VG4d (PL2), VG6d (L2), VG8d (Z2), VG10d (P2), VG13d (D2), VG14d (D3, excluded later), VG25d (Spz2)
Group 3: VG15d (koSpg1), VG16d (koPL3), VG17d(koPL1), VG18d(koPL2), VG19d(koLZp1), VG21d(koLZp2)

* Group 1 = WT biological replicate 1
* Group 2 = WT biological replicate 2
* Group 3 = Mael-/- biological replicates 1 and 2

```
R
library(bsseq)
```

Import DNA methylation

```
#Import group 1

VG1d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/FACS_BSSEQ2014_Rep1/Sample_VG1/rawalign/methextract/
VG1d.merged_CpG_evidence.cov", sampleNames = "wtSpg1", rmZeroCov = T)
#skip to last in the group
VG24d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/methylation/map/VG24/methextract/VG24d.merged_CpG_ev
idence.cov", sampleNames = "wtSpz1", rmZeroCov = T)

#Import group 2

VG2d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/FACS_BSSEQ2014_Rep2/Sample_VG2/rawalign/methextract/
VG2d.merged_CpG_evidence.cov", sampleNames = "wtSpg2", rmZeroCov = T)
#skip to
VG25d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/methylation/map/VG25/methextract/VG25d.merged_CpG_ev
idence.cov", sampleNames = "wtSpz2", rmZeroCov = T)

#Import group 3

VG15d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/methylation/map/VG15/methextract/VG15d.merged_CpG_ev
idence.cov", sampleNames = "koSpg1", rmZeroCov = T)
#skip to
VG21d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/methylation/map/VG21/methextract/VG21d.merged_CpG_ev
idence.cov", sampleNames = "koLZP2", rmZeroCov = T)
```

Obtain "raw" CpG coverage and methylation levels

***Summary for all samples can be found in **Supplementary Table 3**

```
#Calculate mean coverage
colMeans(getCoverage(VG1d))
#[1] 3.673771

#Calculate cytosines examined
sum(rowSums(getCoverage(VG1d)))
#[1] 70741691

#Calculate mean "raw" methylation levels
colMeans(getMeth(VG1d, type = "raw"))
#[1] 0.810014

#Repeat for all samples
```

Extract "raw" DNA methylation and coverage for CpGs covered in common by all samples in a group (common CpGs).
Note, common CpGs were calculated per group (see above)

```
### Extract "raw" DNA methylation and coverage for common CpGs

#Combine and save
VG_WTrep1_full <- combineList(list(VG1d, VG3d, VG5d, VG7d, VG911d, VG12d,
VG24d))
save(VG_WTrep1_full, file = "VG_WTrep1_full.rda")

VG_WTrep1_full
# An object of type 'BSseq' with
#   21046103 methylation loci
#   7 samples
# has not been smoothed

#Examine how many CpGs are covered in common between all samples
sum(rowSums(getCoverage(VG_WTrep1_full) >=1) == 7)
#[1] 13882854

#Extract and export coverage file of all CpGs in a file
VG_WTrep1_full.cov <- getCoverage(VG_WTrep1_full)
save(VG_WTrep1_full.cov, file = "VG_WTrep1_full.cov")

#Select and export coverage file of common CpGs
VG_WTrep1_full_keepLoci <- which(rowSums(VG_WTrep1_full.cov[ , ] >=1) == 7)
save(VG_WTrep1_full_keepLoci, file = "VG_WTrep1_full_keepLoci.rda")

#subset raw methylation data for common CpGs
VG_WTrep1_full_raw_fit_Loci <- VG_WTrep1_full[VG_WTrep1_full_keepLoci, ]
save(VG_WTrep1_full_raw_fit_Loci, file = "VG_WTrep1_full_raw_fit_Loci.rda")

#Check that number of CpGs matches the number of common CpGs
VG_WTrep1_full_raw_fit_Loci
# An object of type 'BSseq' with
#   13882854 methylation loci
#   7 samples
# has not been smoothed

#Repeat for other groups.
```

Summarize raw Methylation and Plot "raw" CpG methylation across common CpGs
***Summary for all samples can be found in **Supplementary Table 5**

```
x <- getMeth(VG_WTrep1_full_raw_fit_Loci, type = "raw")
```

```
colnames(x) <- c("Spg1","PL1", "L1", "Z1", "P1", "D1", "Spz1")

y <- getMeth(VG_WTrep2_full_raw_fit_Loci[, c(1:6,8)], type = "raw")
colnames(y) <- c("Spg2","PL2", "L2", "Z2", "P2", "D2", "Spz2")

head(x)
# Spg1 PL1        L1        Z1 P1 D1 Spz1
# [1,]    1 1.0 0.8000000 0.7500000  1  1    0
# [2,]    1 1.0 0.2000000 0.8000000  1  1    1
# [3,]    1 1.0 0.0000000 1.0000000  1  1    1
# [4,]    1 1.0 0.6666667 1.0000000  1  1    1
# [5,]    1 1.0 1.0000000 0.7142857  1  1    1
# [6,]    1 0.5 0.6000000 1.0000000  1  1    1

head(y)

#Combine biological replicates into one column and rename
bind <- rbind(x,y)
colnames(bind) <- c("Spg","PL", "L", "Z", "P", "D", "Spz")

#Summarize methylation in bins
binsbind <- rep( 1:(nrow(bind)/500), each=500 )
testbind <- as.matrix( aggregate( bind, list(binsbind[1:nrow(bind)]),
na.rm=TRUE, mean ) )

summary(testbind)
#     Group.1           Spg               PL                L
#  Min.   :    1   Min.   :0.08282   Min.   :0.05611   Min.   :0.07197
#  1st Qu.:14478   1st Qu.:0.78422   1st Qu.:0.65979   1st Qu.:0.71665
#  Median :28954   Median :0.84827   Median :0.72189   Median :0.77411
#  Mean   :28954   Mean   :0.82771   Mean   :0.70979   Mean   :0.76059
#  3rd Qu.:43431   3rd Qu.:0.89472   3rd Qu.:0.77569   3rd Qu.:0.82345
#  Max.   :57908   Max.   :0.97644   Max.   :0.92996   Max.   :0.94939
#        Z                 P                 D                Spz
#  Min.   :0.06248   Min.   :0.07493   Min.   :0.07474   Min.   :0.0774
#  1st Qu.:0.74953   1st Qu.:0.79943   1st Qu.:0.80631   1st Qu.:0.8077
#  Median :0.81008   Median :0.86528   Median :0.87270   Median :0.8744
#  Mean   :0.79211   Mean   :0.84371   Mean   :0.85111   Mean   :0.8527
#  3rd Qu.:0.85672   3rd Qu.:0.91486   3rd Qu.:0.92390   3rd Qu.:0.9248
#  Max.   :0.95725   Max.   :0.97877   Max.   :0.97852   Max.   :0.9847

write.table(testbind, file = "VG_WTreps_full_RawMeth_Bin500.txt", sep = "\t",
row.names=FALSE, col.names=FALSE)
```

Plot "raw" CpG methylation across common CpGs

***Summary for all samples can be found in Figure 2A

```
Reps = read.table("VG_WTreps_full_RawMeth_Bin500.txt")

png(filename = "boxplot.WTReps_RawMeth_full_Bin500_colors2.png", width = 800,
height = 500, units = "px")
par(mar=c(4.1, 4.1, 4.1, 2), font.axis=2, font.lab=2, cex.lab = 1.8, cex.axis =
2.2, las=0.5)
boxplot(Reps[, c(2:8)], range =0, ylim=c(0,1), axes=T,
col=c("blue","red", "green", "orange", "dark violet", "sky blue", "grey"),
names= c("Spg", "PL", "L", "Z", "P", "D", "Spz"),
ylab='Mean CpG Methylation Levels')
dev.off()
```

Protocol 3. Analysis of DMRs with *bsseq* package

This Protocol was written and executed by Valeriya Gaysinskaya, with guidance from Kasper Daniel Hansen.
The Protocol was adopted from “Analyzing WGBS with the bases package" by Kasper Daniel Hansen.

This document contains a pipeline for extracing and analyzing Differentially Methylated Regions (DMRs) from WGBS data with Bsseq package, pre-processed by Bismark. The document focuses on analysis-related tasks and questions.

Quick References:
Bsseq (http://bioconductor.org/packages/release/bioc/html/bsseq)

R version 3.0.1 (2013-05-16) was used
Bsseq version 0.10.0 was used

Sample IDs (here, samples were grouped based on their Illumina Sequencing date):
Group 1: VG1d (Spg1), VG3d (PL1), VG5d (L1), VG7d (Z1), VG911d (P1), VG12d (D1)
Group 2: VG2d (Spg2), VG4d (PL2), VG6d (L2), VG8d (Z2), VG10d (P2), VG13d (D2), VG14d (D3, excluded)
Group 3: VG15d (koSpg1), VG16d (koPL3), VG17d(koPL1), VG18d(koPL2), VG19d(koLZp1), VG21d(koLZp2), VG24d (Spz1), VG25d (Spz2)

* Group 1 = WT biological replicate 1
* Group 2 = WT biological replicate 2
* Group 3 = Mael-/- biological replicates 1 and 2

PART I: IMPORT, COMBINE and SPLIT.

In Part I, individual samples are grouped and then split again, to account for most CpGs.

```{r}
# Readfiles and combine
R
library(bsseq)

#Import datasets

#Import group 1

VG1d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/FACS_BSSEQ2014_Rep1/Sample_VG1/rawalign/methextract/
VG1d.merged_CpG_evidence.cov", sampleNames = "wtSpg1", rmZeroCov = T)
#skip to last in the group
VG12d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/FACS_BSSEQ2014_Rep1/Sample_VG12/rawalign/methextract
/VG12d.merged_CpG_evidence.cov", sampleNames = "wtD1", rmZeroCov = T)

# Combine and save
VG_WTrep1 <- combineList(list(VG1d, VG3d, VG5d, VG7d, VG911d, VG12d))
save(VG_WTrep1, file = "VG_WTrep1.rda")


#Import group 2

VG2d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/FACS_BSSEQ2014_Rep2/Sample_VG2/rawalign/methextract/
VG2d.merged_CpG_evidence.cov", sampleNames = "wtSpg2", rmZeroCov = T)
#skip to last in the group
```

```
VG14d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/methylation/map/VG14/methextract/VG14d.merged_CpG_ev
idence.cov", sampleNames = "wtD3", rmZeroCov = T)

#Combine and save
VG_WTrep2 <- combineList(list(VG2d, VG4d, VG6d, VG8d, VG10d, VG13d, VG14d))
save(VG_WTrep2, file = "VG_WTrep2.rda")

#Import group 3

VG15d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/methylation/map/VG15/methextract/VG15d.merged_CpG_ev
idence.cov", sampleNames = "koSpg1", rmZeroCov = T)
#skip to last in the group
VG25d <- read.bismark(files =
"/mnt/sequence/gaysinskaya/methylation/map/VG25/methextract/VG25d.merged_CpG_ev
idence.cov", sampleNames = "wtSP2", rmZeroCov = T)

###Note, that koSpg2 is actually a koPL3 sample!!!!

#Combine and save
VG_bsseq3 <- combineList(list(VG15d, VG16d, VG17d, VG18d, VG19d, VG21d, VG24d,
VG25d))
save(VG_bsseq3, file = "VG_bsseq3.rda")

```


Summarize CpG coverage, DNA methylation and total number of CpGs covered
(mappable CpGs)
***The results can be found in Supplementary Tables xxxx

```{r, echo=FALSE}

#Explore Rep1
VG1d
# An object of type 'BSseq' with
#   19255878 methylation loci
#   1 samples
# has not been smoothed

#skip to last

VG12d
# An object of type 'BSseq' with
#   20491322 methylation loci
#   1 samples
# has not been smoothed

###
#Average Cov
# Total # of CpGs
#Average Meth

#Extract CpG coverage and total number of covered CpGs
colMeans(getCoverage(VG1d))
#[1] 3.673771
sum(rowSums(getCoverage(VG1d)))
#[1] 70741691
colMeans(getMeth(VG1d, type = "raw"))
#[1] 0.810014

#skip to the last
```

```
colMeans(getCoverage(VG3d))
#[1] 3.436042
sum(rowSums(getCoverage(VG3d)))
#[1] 64944450
colMeans(getMeth(VG3d, type = "raw"))
#[1] 0.6673618
```

From the above evaluation, we can see that each sample was covered at a slightly different depth, covering a slightly different number of CpGs. By combining samples into one group above we accounted for more genomic CpGs. Where the CpG coverage is zero, combining samples resultit in the non-covered CpG having an 'NA' instead.

```
```

For DMR analysis it is important to account for most CpGs spatially. Thus, even if a particular CpG is not available for analysis and is 'NA', leaving that unavailable CpG is important for the Bsseq package to know that there is a CpG there. Thus, instead of using individual samples that were originally imported into Bsseq, it is recommended, instead, to split the combined file into individual samples (thus accounting for more CpG sites).

Split samples from combined into individual.

```{r}

#Split samples from Group 1

VG1c <- VG_WTrep1[, 1]
save(VG1c, file = "VG1c.rda")
#skip to last
VG12c <- VG_WTrep1[, 6]
save(VG12c, file = "VG12c.rda")

#Split samples from Group 2

VG2c <- VG_WTrep2[, 1]
save(VG2c, file = "VG2c.rda")
#skip to last
VG14c <- VG_WTrep2[, 7]
save(VG14c, file = "VG14c.rda")

#Split samples from Group 3

VG15c <- VG_bsseq3[, 1]
save(VG15c, file = "VG15c.rda")
#skip to last
VG25c <- VG_bsseq3[, 8]
save(VG25c, file = "VG25c.rda")

```
```

**Part II. PREPARE SAMPLES FOR PAIRWISE DMR ANALYSIS.**

In Part II, biological replicates of a pairwise comparison are combined for a 2x2 analysis.
'Random' chromosomes are eliminated, because they are very short, and lead to false interpretations.
An example is used here involving Spg and PL samples.

WT SpgPL
```{r}

```
#Check column names for correct samples
> colnames(VG1c)
#[1] "wtSpg1"
> colnames(VG3c)
#[1] "wtPL1"
> colnames(VG2c)
#[1] "wtSpg2"
> colnames(VG4c)
#[1] "wtPL2"

#Combine into one file
WTSpgPL <- combineList(list(VG1c, VG3c, VG2c, VG4c))
save(WTSpgPL, file = "WTSpgPL.rda")

WTSpgPL
# An object of type 'BSseq' with
#   21140248 methylation loci
#   4 samples
# has not been smoothed

head(getCoverage(WTSpgPL))

#      [,1] [,2] [,3] [,4]
# [1,]    6    3    4    8
# [2,]    3    3    1    4
# [3,]    3    1    0    1
# [4,]    4    0    2    7
# [5,]    5    1    3    4
# [6,]    6    1    2    4

#Examine the number of CpG sites covered by all 4 samples at least once
sum(rowSums(getCoverage(WTSpgPL) >=1) == 4)
#[1] 16838779

# Get a summary of CpG coverage
summary(getCoverage(WTSpgPL))
#        V1                 V2                  V3                 V4
#  Min.   :    0.00   Min.   :    0.00   Min.   :    0.0   Min.   :    0.00
#  1st Qu.:    2.00   1st Qu.:    1.00   1st Qu.:    2.0   1st Qu.:    3.00
#  Median :    3.00   Median :    3.00   Median :    4.0   Median :    5.00
#  Mean   :    3.35   Mean   :    3.07   Mean   :    4.4   Mean   :    5.02
#  3rd Qu.:    4.00   3rd Qu.:    4.00   3rd Qu.:    6.0   3rd Qu.:    7.00
#  Max.   :72672.00   Max.   :63392.00   Max.   :73959.0   Max.   :67439.00


#Examine Random Chromosomes and decide which to eliminate

seqnames(WTSpgPL)
# factor-Rle of length 21140248 with 35 runs
#   Lengths:      1452905      1090019      1154052 ...        13798
186820
#   Values :         chr1        chr10        chr11 ...         chrY
chrY_random
# Levels(35): chr1 chr10 chr11 chr12 chr13 ... chrX chrX_random chrY
chrY_random

#Examine chromosome names
seqnames(WTSpgPL)@values
#  [1] chr1         chr10        chr11        chr12        chr13
#  [6] chr13_random chr14        chr15        chr16        chr16_random
# [11] chr17        chr17_random chr18        chr19        chr1_random
# [16] chr2         chr3         chr3_random  chr4         chr4_random
# [21] chr5         chr5_random  chr6         chr7         chr7_random
```

```
# [26] chr8         chr8_random  chr9         chr9_random  chrM
# [31] chrUn_random chrX         chrX_random  chrY         chrY_random
# 35 Levels: chr1 chr10 chr11 chr12 chr13 chr13_random chr14 chr15 ...
chrY_random

#Examine lengths of chromosomes
seqnames(WTSpgPL)@lengths
# [1] 1452905 1090019 1154052  935939  963041     755  910348  860439  744977
# [10]      19  835358     729  714134  552845    7768 1495101 1173611      22
# [19] 1299246    1245 1329239     204 1147014 1187167     450 1111831    3337
# [28] 1057470    2491     273   16770  888289    2542   13798  186820

#Connect chr name with chr length for easy visualization and subsequent double
checking of final length
cbind( seqnames(WTSpgPL)@lengths, seqlevels(WTSpgPL))
#      [,1]      [,2]
#  [1,] "1452905" "chr1"
#  [2,] "1090019" "chr10"
#  [3,] "1154052" "chr11"
#  [4,] "935939"  "chr12"
#  [5,] "963041"  "chr13"
#  [6,] "755"     "chr13_random"
#  [7,] "910348"  "chr14"
#  [8,] "860439"  "chr15"
#  [9,] "744977"  "chr16"
# [10,] "19"      "chr16_random"
# [11,] "835358"  "chr17"
# [12,] "729"     "chr17_random"
# [13,] "714134"  "chr18"
# [14,] "552845"  "chr19"
# [15,] "7768"    "chr1_random"
# [16,] "1495101" "chr2"
# [17,] "1173611" "chr3"
# [18,] "22"      "chr3_random"
# [19,] "1299246" "chr4"
# [20,] "1245"    "chr4_random"
# [21,] "1329239" "chr5"
# [22,] "204"     "chr5_random"
# [23,] "1147014" "chr6"
# [24,] "1187167" "chr7"
# [25,] "450"     "chr7_random"
# [26,] "1111831" "chr8"
# [27,] "3337"    "chr8_random"
# [28,] "1057470" "chr9"
# [29,] "2491"    "chr9_random"
# [30,] "273"     "chrM"
# [31,] "16770"   "chrUn_random"
# [32,] "888289"  "chrX"
# [33,] "2542"    "chrX_random"
# [34,] "13798"   "chrY"
# [35,] "186820"  "chrY_random"

#Drop the chromosomes listed below and retain the rest:
#Also Retain chr1_random only: largest chromosome in mouse genome, so even
random contig presumably contains enough information
#Also, retain "chrX_random" and "chrY_random"

# [6,] "755"     "chr13_random"
# [10,] "19"      "chr16_random"
# [12,] "729"     "chr17_random"
# [18,] "22"      "chr3_random"
# [20,] "1245"    "chr4_random"
# [22,] "204"     "chr5_random"
# [25,] "450"     "chr7_random"
```

```
# [27,] "3337"    "chr8_random"
# [29,] "2491"    "chr9_random"
# [30,] "273"     "chrM"
# [31,] "16770"   "chrUn_random"

Eliminate unwanted random chromosomes

#force - TRUE. Force dropping sequence levels currently in use.
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr13_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr16_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr17_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr3_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr4_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr5_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr7_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr8_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chr9_random")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chrM")]
# seqlevels(WTSpgPL, force=TRUE) <- seqlevels(WTSpgPL)[seqlevels(WTSpgPL) !=
c("chrUn_random")]

WTSpgPL_rm <- WTSpgPL
save(WTSpgPL_rm, file = "WTSpgPL_rm.rda")
rm WTSpgPL

#Explore Bsseq object after removing unwanted chr
WTSpgPL_rm
# An object of type 'BSseq' with
#   21113953 methylation loci
#   4 samples
# has not been smoothed

summary(getCoverage(WTSpgPL_rm))
#        V1                V2                V3                V4
#  Min.   :    0.00   Min.   :    0.00   Min.   :    0.00   Min.   :    0.00
#  1st Qu.:    2.00   1st Qu.:    1.00   1st Qu.:    2.00   1st Qu.:    3.00
#  Median :    3.00   Median :    3.00   Median :    4.00   Median :    5.00
#  Mean   :    3.34   Mean   :    3.07   Mean   :    4.39   Mean   :    5.02
#  3rd Qu.:    4.00   3rd Qu.:    4.00   3rd Qu.:    6.00   3rd Qu.:    7.00
#  Max.   :72672.00   Max.   :63392.00   Max.   :73959.00   Max.   :67439.00

```

Part III. SMOOTHING

Previously we determined that there are zero coverage CpGs in every sample, and that they can be different for each sample. In Part III, the DNA methylation level in a genomic regions is smoothed, using local likelihood estimation, for each sample. As a result, after smoothiing, every CpG in the objectect has an estimated methylation value, based on information from covered CpGs.

Smoothing was done by chromosome using large DMR settings as outlined in Kasper D Hansen et al, Genome Biology 2012).

For smoothing in large blocks, a window size of 500 CpGs or 20,000 basepairs, whichever was larger, was used for smoothing.
An example of smoothing for Spg and PL samples is shown below.

```{r}
load("WTSpgPL_rm.rda")
WTSpgPL_LGsmooth <- BSmooth(WTSpgPL_rm, parallelBy = c("chromosome"), ns = 500, h = 20000, maxGap = 10^8, mc.cores = 1, verbose = T)
save(WTSpgPL_LGsmooth, file = "WTSpgPL_LGsmooth.rda")

#[BSmooth] smoothing done in 3847.1 sec

An object of type 'BSseq' with
#   21113953 methylation loci
#   4 samples
# has been smoothed with
#   BSmooth (ns = 500, h = 20000, maxGap = 100000000)

#The coverage should not change
summary(getCoverage(WTSpgPL_LGsmooth))
#        V1                 V2                 V3                 V4
#  Min.   :    0.00   Min.   :    0.00   Min.   :    0.00   Min.   :    0.00
#  1st Qu.:    2.00   1st Qu.:    1.00   1st Qu.:    2.00   1st Qu.:    3.00
#  Median :    3.00   Median :    3.00   Median :    4.00   Median :    5.00
#  Mean   :    3.34   Mean   :    3.07   Mean   :    4.39   Mean   :    5.02
#  3rd Qu.:    4.00   3rd Qu.:    4.00   3rd Qu.:    6.00   3rd Qu.:    7.00
#  Max.   :72672.00   Max.   :63392.00   Max.   :73959.00   Max.   :67439.00

#The raw DNA methylation values have been smoothed
summary(getMeth(WTSpgPL_LGsmooth, type = "smooth"))
#    wtSpg1           wtPL1              wtSpg2             wtPL2
#  Min.   :0.0000   Min.   :0.0000018   Min.   :0.0000895   Min.   :0.01287
#  1st Qu.:0.7437   1st Qu.:0.6090292   1st Qu.:0.7305578   1st Qu.:0.65196
#  Median :0.8737   Median :0.7081948   Median :0.8521176   Median :0.76311
#  Mean   :0.8099   Mean   :0.6725245   Mean   :0.7906918   Mean   :0.71570
#  3rd Qu.:0.9297   3rd Qu.:0.7636447   3rd Qu.:0.9024897   3rd Qu.:0.81378
#  Max.   :1.0000   Max.   :1.0000000   Max.   :1.0000000   Max.   :1.00000

#The raw DNA methylation profile, before smoothing, contains many NA values
summary(getMeth(WTSpgPL_LGsmooth, type = "raw"))
#        V1               V2               V3               V4
#  Min.   :0.0       Min.   :0.0       Min.   :0.0       Min.   :0.0
#  1st Qu.:0.8       1st Qu.:0.5       1st Qu.:0.7       1st Qu.:0.6
#  Median :1.0       Median :0.8       Median :1.0       Median :0.8
#  Mean   :0.8       Mean   :0.7       Mean   :0.8       Mean   :0.7
#  3rd Qu.:1.0       3rd Qu.:1.0       3rd Qu.:1.0       3rd Qu.:1.0
#  Max.   :1.0       Max.   :1.0       Max.   :1.0       Max.   :1.0
#  NA's   :1876411   NA's   :2230001   NA's   :1165627   NA's   :1099637

```

Part IV. COMPUTING t-statistics

A. Before computing t-statistics, CpGs with no coverage are removed, otherwise, most likely resulting in false positives. For this analysis we will only keep CpGs where all four samples in a pairwise comparison are covered at least once.

```{r}

#Extract covrage into a separate matrix
WTSpgPL_LGsmooth.cov <- getCoverage(WTSpgPL_LGsmooth)
save(WTSpgPL_LGsmooth.cov, file = "WTSpgPL_LGsmooth.cov")
```

```
#Check the number of CpGs covered by all four samples at least once (common
CpGs)
sum(rowSums(getCoverage(WTSpgPL_LGsmooth[ , ]) >=1) == 4)
#[1] 16825193

#Select for coverage of common CpGs
WTSpgPL_keepLociLG <- which(rowSums(WTSpgPL_LGsmooth.cov[ , ] >=1) == 4)
save(WTSpgPL_keepLociLG, file = "WTSpgPL_keepLociLG.rda")

#Subset smoothed data with select common loci
WTSpgPL_fit_lociLG <- WTSpgPL_LGsmooth[WTSpgPL_keepLociLG, ]
save(WTSpgPL_fit_lociLG, file = "WTSpgPL_fit_lociLG.rda")
WTSpgPL_fit_lociLG
# An object of type 'BSseq' with
#   16825193 methylation loci
#   4 samples
# has been smoothed with
#   BSmooth (ns = 500, h = 20000, maxGap = 100000000)

summary(getCoverage(WTSpgPL_fit_lociLG))
#       V1                 V2                 V3                V4
#  Min.   :    1.00   Min.   :    1.00   Min.   :    1.0   Min.   :    1.0
#  1st Qu.:    2.00   1st Qu.:    2.00   1st Qu.:    3.0   1st Qu.:    3.0
#  Median :    3.00   Median :    3.00   Median :    4.0   Median :    5.0
#  Mean   :    3.83   Mean   :    3.56   Mean   :    4.9   Mean   :    5.6
#  3rd Qu.:    5.00   3rd Qu.:    4.00   3rd Qu.:    6.0   3rd Qu.:    7.0
#  Max.   :72672.00   Max.   :63392.00   Max.   :73959.0   Max.   :67439.0

summary(getMeth(WTSpgPL_fit_lociLG))
#      wtSpg1            wtPL1            wtSpg2            wtPL2
#  Min.   :0.01288   Min.   :0.0243   Min.   :0.01679   Min.   :0.02175
#  1st Qu.:0.76614   1st Qu.:0.6217   1st Qu.:0.75067   1st Qu.:0.66745
#  Median :0.88229   Median :0.7120   Median :0.85893   Median :0.76765
#  Mean   :0.82250   Mean   :0.6791   Mean   :0.80209   Mean   :0.72377
#  3rd Qu.:0.93245   3rd Qu.:0.7649   3rd Qu.:0.90475   3rd Qu.:0.81513
#  Max.   :1.00000   Max.   :0.9953   Max.   :0.99991   Max.   :0.99946

```

B. T-statistics are formed as the difference in means between group 1 and group 2 (group1 - group2) divided by an estimate of the standard deviation, assuming that the variance in the two groups is paired samples (paired). The standard deviation estimates are then smoothed (using a running mean with a width of k) and thresholded (using qSd which sets the minimum standard deviation to be the qSd-quantile). The parameters used are as described in Kasper D Hansen et al, Genome Biology 2012, where the standard errors were smoothed using a running mean with a window size of 101 observations, and the threshold for the standard deviation of its 75th percentile was applied.

```{r}
WTSpgPL_fit_lociLG.tstat <- BSmooth.tstat(WTSpgPL_fit_lociLG,
                                          group1 = c(1, 3),
                                           group2 = c(2, 4),
                                           maxGap = 100000000,
                                           local.correct = FALSE,
                                           estimate.var = "paired",
                                           qSd = 0.75,
                                           k = 101,
                                           mc.cores = 1,
                                           verbose = TRUE)

save(WTSpgPL_fit_lociLG.tstat, file = "WTSpgPL_fit_lociLG.tstat.rda")
```

```
WTSpgPL_fit_lociLG.tstat
# An object of type 'BSseqTstat' with
#   16825193 methylation loci
# based on smoothed data:
#   BSmooth (ns = 500, h = 20000, maxGap = 100000000)
# with parameters
#   BSmooth.tstat (local.correct = FALSE, maxGap = 100000000)

summary(WTSpgPL_fit_lociLG.tstat@stats)
#     rawSds           tstat.sd        group2.means      group1.means
#  Min.   :0.00000   Min.   :0.01041   Min.   :0.02519   Min.   :0.01483
#  1st Qu.:0.02917   1st Qu.:0.05028   1st Qu.:0.64527   1st Qu.:0.75853
#  Median :0.05044   Median :0.05153   Median :0.74099   Median :0.87046
#  Mean   :0.05258   Mean   :0.05475   Mean   :0.70145   Mean   :0.81230
#  3rd Qu.:0.07202   3rd Qu.:0.05469   3rd Qu.:0.78849   3rd Qu.:0.91830
#  Max.   :0.37692   Max.   :0.24981   Max.   :0.99736   Max.   :0.99996
#      tstat
#  Min.   :-5.648
#  1st Qu.: 1.403
#  Median : 2.169
#  Mean   : 2.046
#  3rd Qu.: 2.807
#  Max.   : 9.569


```
```

Part V. FINDING DMRs

Once t-statistics have been computed, we can compute differentially methylated regions (DMRs) by thresholding the t-statistics. Here we use a quintile-based cutoff of 0.01 and 0.99 of the t-statistics (so, for most of the entire genome).

```{r}

#Define ALL LG (LARGE) DMRs

WTSpgPL_LGdmr0 <- dmrFinder(WTSpgPL_fit_lociLG.tstat, maxGap = 10000, stat =
"tstat", cutoff = c(-2, 2), verbose = TRUE)
save(WTSpgPL_LGdmr0 , file = "WTSpgPL_LGdmr0.rda")

# To detect large DMR blocks, specify that a block needs to be at least 5kb in
size
WTSpgPL_LGdmrs <- WTSpgPL_LGdmr0[WTSpgPL_LGdmr0$width >= 5000, ]
save(WTSpgPL_LGdmrs , file = "WTSpgPL_LGdmrs.rda")

#Explore

summary(WTSpgPL_LGdmrs)
#     idxEnd            cluster             n               width
#  Min.   :    6566   Min.   :   1.0   Min.   :    2.0   Min.   :   5000
#  1st Qu.: 4274762   1st Qu.:  70.0   1st Qu.:  129.0   1st Qu.:  16786
#  Median : 8497647   Median : 182.0   Median :  257.0   Median :  35223
#  Mean   : 8587683   Mean   : 223.1   Mean   :  495.9   Mean   :  72252
#  3rd Qu.:12941212   3rd Qu.: 343.0   3rd Qu.:  563.0   3rd Qu.:  80898
#  Max.   :16825186   Max.   :1631.0   Max.   :12215.0   Max.   :1972045
#
#    invdensity        areaStat          maxStat          meanDiff
#  Min.   :  39.3   Min.   : -495.6   Min.   :-2.894   Min.   :-0.2498
#  1st Qu.: 101.4   1st Qu.:  299.5   1st Qu.: 2.414   1st Qu.: 0.1176
#  Median : 129.2   Median :  653.2   Median : 2.883   Median : 0.1342
#  Mean   : 161.2   Mean   : 1374.0   Mean   : 2.961   Mean   : 0.1356
#  3rd Qu.: 182.8   3rd Qu.: 1526.5   3rd Qu.: 3.447   3rd Qu.: 0.1535
```
```

```
#  Max.   :4821.5   Max.   :39755.2   Max.   : 9.569   Max.   : 0.3337
#
#   group1.mean      group2.mean        tstat.sd        direction
#  Min.   :0.2043   Min.   :0.1818   Min.   :0.01244   Length:19019
#  1st Qu.:0.8175   1st Qu.:0.6760   1st Qu.:0.05078   Class :character
#  Median :0.8708   Median :0.7284   Median :0.05263   Mode  :character
#  Mean   :0.8484   Mean   :0.7128   Mean   :0.05384
#  3rd Qu.:0.9041   3rd Qu.:0.7679   3rd Qu.:0.05545
#  Max.   :0.9644   Max.   :0.8640   Max.   :0.10565

#Number of CpG loci
sum(WTSpgPL_LGdmrs$n)
#[1] 9432366

#Genomic size
sum(WTSpgPL_LGdmrs$width)
#[1]  1374153075


#Total # of CpG loci
x <- WTSpgPL_LGdmrs[WTSpgPL_LGdmrs$direction == "hyper", ]
y <- WTSpgPL_LGdmrs[WTSpgPL_LGdmrs$direction == "hypo", ]

#Summarize, e.g., Hypermethylated DMRs (hyper in Spg, hypo in PL)
summary(x)
#    idxEnd            cluster           n               width
#  Min.   :    6566   Min.   :   1   Min.   :    2.0   Min.   :   5000
#  1st Qu.: 4269070   1st Qu.:  70   1st Qu.:  130.0   1st Qu.:  16926
#  Median : 8494976   Median : 182   Median :  258.0   Median :  35440
#  Mean   : 8583517   Mean   : 223   Mean   :  497.4   Mean   :  72479
#  3rd Qu.:12935981   3rd Qu.: 343   3rd Qu.:  565.0   3rd Qu.:  81170
#  Max.   :16825186   Max.   :1631   Max.   :12215.0   Max.   :1972045
#
#    invdensity         areaStat           maxStat         meanDiff
#  Min.   :  40.08   Min.   :    4.18   Min.   :2.003   Min.   :0.02971
#  1st Qu.: 101.52   1st Qu.:  302.26   1st Qu.:2.419   1st Qu.:0.11778
#  Median : 129.28   Median :  655.91   Median :2.886   Median :0.13429
#  Mean   : 161.06   Mean   : 1379.81   Mean   :2.979   Mean   :0.13662
#  3rd Qu.: 182.89   3rd Qu.: 1530.16   3rd Qu.:3.449   3rd Qu.:0.15355
#  Max.   :4821.50   Max.   :39755.16   Max.   :9.569   Max.   :0.33366
#
#   group1.mean      group2.mean        tstat.sd        direction
#  Min.   :0.3310   Min.   :0.1818   Min.   :0.01244   Length:18951
#  1st Qu.:0.8184   1st Qu.:0.6768   1st Qu.:0.05078   Class :character
#  Median :0.8712   Median :0.7287   Median :0.05263   Mode  :character
#  Mean   :0.8500   Mean   :0.7134   Mean   :0.05384
#  3rd Qu.:0.9043   3rd Qu.:0.7680   3rd Qu.:0.05546
#  Max.   :0.9644   Max.   :0.8640   Max.   :0.10565


#Number of CpG loci
sum(x$n)
#[1] 9426532

#Genomic size
sum(x$width)
#[1] 1373552310

#Write out final dmr.txt file
write.table(x=WTSpgPL_LGdmrs, file="WTSpgPL_LGdmrs.txt", sep="\t", row.names=
FALSE, quote=FALSE)
write.table(x=x, file="WTSpgPL_LGdmrs_hyper.txt", sep="\t", row.names= FALSE,
quote=FALSE)
```

```
write.table(x=y, file="WTSpgPL_LGdmrs_hypo.txt", sep="\t", row.names= FALSE,
quote=FALSE)
```

APPENDIX

COMPUTATIONAL SCRIPTS 1-7: Processing of Illumina Bisulfite-seq reads

```
# This pipeline is for the processing of Illumina Bisulfite-Sequencing reads, sequenced on HiSeq2000. The goal is to obtain a file containing CpG methylation evidence (methylation values and coverage data).
## The pipeline uses a subset of samples from sequencing run 1 for demonstration, and includes samples 1,3,5,7,9 and 12 corresponding to wild-type samples Spg1, PL1, L1, Z1, P1 and D1, respectively.



### Script1: Concatenate all fastq.gz files and run FASTQ analysis

#!/bin/bash

# This script is "doGenerateFastq-WT-Rep1.sh"
# This file concatenates all fastq.gz files for a particular sample and runs FASTQ analysis for wild-type (WT) samples of sequencing run 1 (sequenced on HiSeq2000)

# view FASTQ directory and examine sample folders
       #Sample_VG1  Sample_VG12  Sample_VG3  Sample_VG5  Sample_VG7  Sample_VG9

FASTQ_DIR=/mnt/sequence/FastQ/141002_SN375_0277_BC5LM0ACXX/FastQ/Project_AB-valeriyagaysinkaya

#SAMPLE="1 3 5 7 9 12"

mkdir -p rawdata
cd rawdata

###For all samples
for i in $SAMPLE
do
echo "Concatenating VG$i (Read 1)"
  cat $FASTQ_DIR/Sample_VG$i/*_R1_*.fastq.gz > VG$i\_R1.fastq.gz
  echo "Concatenating VG$i (Read 2)"
  cat $FASTQ_DIR/Sample_VG$i/*_R2_*.fastq.gz > VG$i\_R2.fastq.gz
  fastqc VG$i\_R1.fastq.gz &
  fastqc VG$i\_R2.fastq.gz &
done
                                  *************



### Script2: Concatenate different samples, if necessary (e.g., to increase coverage)

#!/bin/bash


FASTQ_DIR=/mnt/sequence/gaysinskaya/methylation/rawdata

#mkdir -p rawdata
#cd rawdata

 # echo "Concatenating Sample VG9 with Sample VG11 (Read 1)"
 # cat $FASTQ_DIR/VG9_R1.fastq.gz $FASTQ_DIR/VG11_R1.fastq.gz > VG911\_R1.fastq.gz
 # echo "Concatenating Sample VG9 with Sample VG11 (Read 2)"
 # cat $FASTQ_DIR/VG9_R2.fastq.gz $FASTQ_DIR/VG11_R2.fastq.gz > VG911\_R2.fastq.gz
 # fastqc VG911_R1.fastq.gz &
 # fastqc VG911_R2.fastq.gz &
```

```
mkdir -p rawdata
cd rawdata

  echo "Concatenating Sample VG19 with Sample VG20 (Read 1)"
  cat $FASTQ_DIR/VG19_R1.fastq.gz $FASTQ_DIR/VG20_R1.fastq.gz >
VG1920\_R1.fastq.gz
  echo "Concatenating Sample VG19 with Sample VG20 (Read 2)"
  cat $FASTQ_DIR/VG19_R2.fastq.gz $FASTQ_DIR/VG20_R2.fastq.gz >
VG1920\_R2.fastq.gz
```

****

Script3: Align bisulfite-sequencing samples to mouse mm9 genome using Bismark program

```
#!/bin/bash

# This script is "doBismarkMap_Rep1.sh"
# This file aligns reads to mouse mm9 genome using Bismark, for samples of
sequencing run 1. Note, prior Bismark genome preparation step is required (see
Bismark documentation).
# The alignment was performed using Bismark Version: v0.13.0 (Last edited on 21
July 2014.)

FASTQ_DIR=/mnt/sequence/gaysinskaya/methylation/rawdata
BISMARK_GENOME=/mnt/sequence/gaysinskaya/Bismark_bt2mm9
SAMPLE="1 3 5 7 9 12 20"

Bismark align all samples
for i in $SAMPLE
do
  mkdir -p map
  cd map
  mkdir -p VG$i
  cd VG$i
  echo "Running Bismark on VG$i"
       bismark -q -p 4 --bowtie2 --bam --score_min L,0,-0.4 $BISMARK_GENOME -1
$FASTQ_DIR/VG$i\_R1.fastq.gz -2 $FASTQ_DIR/VG$i\_R

done
```

****

Script4: De-duplicate bisulfite-sequencing reads using Bismark program

```
#!/bin/bash

#This script is "dedup_Rep1.sh" and is used to deduplicate reads in the .bam
file
#Deduplication was performed using a Bismark script last modified on Oct 23,
2013

BAM_DIR=/mnt/sequence/gaysinskaya/methylation/map
SAMPLE="1 3 5 7 9 12 "

# Deduplicate all samples
#If want to deduplicate one sample only (e.g., sample 1), replace "for i in
$SAMPLE" with "for i in 1".

for i in $SAMPLE
```

```
do
  cd $BAM_DIR
  echo "Deduplicating  subset VG$i"
  nohup deduplicate_bismark --paired --bam $BAM_DIR/VG$i/VG$i\_*.bam >
nohupdedup_VG$i&
done
                          **********
```

Script5: Perform M-bias on deduplicated .bam files

```
#!/bin/bash

#This script is to perform M-bias (perform only on deduplicated .bam files)
#The output allows for determination of methylation-bias (M-bias) and
subsequent trimming of biased reads.

#M-bias was performed using Bismark Extractor Version: v0.14.3_devel (last
modified on 22 April 2015.)

DEDUP_DIR=/mnt/sequence/gaysinskaya/methylation/map
SAMPLE="1 3 5 7 9 12 20"

#Run M-bias on all samples in SAMPLE

for i in SAMPLE
do
  cd $DEDUP_DIR/VG$i
  echo "Running M-bias only on VG$i"
  nohup  bismark_methylation_extractorV14_3.pl --multicore 2 --mbias_only --
paired-end $DEDUP_DIR/VG$i/VG$i\_*.deduplicated.bam > nohupMbias_VG$i
done
                          **********
```

Script6: Perform methylation extraction and elimination of M-bias

```
#!/bin/bash

#This script is "dodoMethylExtract.sh" and it performs methylation extraction
and elimination of M-bias, simultaneously.
#Specifically, after examination of M-bias in previous step, this script
extracts methylation information ignoring biased CpGs. The biased CpGs included
6 nt at 5' of read 1, 10 nt at 5' of read 2, 1 nt at 3' of read 1 and read 2.

DEDUP_DIR=/mnt/sequence/gaysinskaya/methylation/map
SAMPLE1="1 3 5 7 9 12 20"

for i in $SAMPLE1
 do
   cd $DEDUP_DIR/VG$i
   mkdir -p methextract

  echo "Extracting methylation from VG$i"
  nohup  bismark_methylation_extractorV14_3.pl --multicore 2 --gzip --paired-
end --no_overlap --report --ignore 6 --ignore_r2 10 --ignore_3prime 1 --
ignore_3prime_r2 1 -o methextract --comprehensive
$DEDUP_DIR/VG$i/VG$i\_*.deduplicated.bam > nohupextract_VG$i
done

#Output notes:
```

```
#Results file containing methylation information for C in CpG context will be
written to
methextract/CpG_context_VG1_R1.fastq.gz_bismark_bt2_pe.deduplicated.txt.gz
#Results files containing methylation information for C in CHG context and C in
CHH context will also be written out.


# view CpG methylation file
#head
CpG_context_VG1_IDX5WTSPG1_L001_R1ALL.fastq.gz_bismark_bt2_pe.deduplicated.txt,
where Z stands for methylated and z for unmethylated

#HWI-ST375:277:C5LM0ACXX:1:1101:1407:2118_1:N:0:     +      chr1   92624580
       Z
#HWI-ST375:277:C5LM0ACXX:1:1101:1407:2118_1:N:0:     +      chr1   92624623
       Z
#HWI-ST375:277:C5LM0ACXX:1:1101:1407:2118_1:N:0:     +      chr1   92624767
       Z
#HWI-ST375:277:C5LM0ACXX:1:1101:1326:2236_1:N:0:     +      chr10  30281603
       Z
#HWI-ST375:277:C5LM0ACXX:1:1101:1631:2039_1:N:0:     +      chr4   73920485
       Z
#HWI-ST375:277:C5LM0ACXX:1:1101:1631:2039_1:N:0:     +      chr4   73920346
       Z
#HWI-ST375:277:C5LM0ACXX:1:1101:1620:2081_1:N:0:     -      chrX   104373998
       z
```

****

```
### Script7: Merge CpG methylation for both strands
# The output file is a final file used for DNA methylation analysis.


#!/bin/bash

# This script is "CpG_ExtractMerge.sh"
# This script extracts strand-specific coverage (using bismark2bedgGraph
script) followed by strand-merging (using coverage2cytosine script)
#Working directory = /mnt/sequence/gaysinskaya/methylation

CpG_context=/mnt/sequence/gaysinskaya/methylation/map
mm9Genome=/mnt/sequence/gaysinskaya/Bismark_bt2mm9

SAMPLE1="1 3 5 7 9 12"

for i in $SAMPLE1
 do
    cd $CpG_context/VG$i/methextract
    echo "Extracting strand-specific CpG coverage from Sample VG$i"
    nohup time bismark2bedGraph -o VG$i CpG_context_VG$i*.deduplicated.txt.gz >
nohupcov
    echo "Merging strand-specific CpG coverage from Sample VG$i"
    nohup time coverage2cytosine -o VG$i\d --merge_CpG --genome_folder
$mm9Genome VG$i\.bismark.cov > nohupcov2
done

 # view FINAL CpG methylation file
 ## head VG1d.merged_CpG_evidence.cov
 ### column ids: chr/CpG start/CpG end/% methyllation/# CpG evidence for
methylated CpG/# CpG evidence for unmethylated CpG

#chr1  3000574      3000575      100.000000   6     0
```

```
#chr1  3000726      3000727      100.000000   3      0
#chr1  3000901      3000902      100.000000   3      0
#chr1  3001346      3001347      100.000000   4      0
#chr1  3001394      3001395      100.000000   5      0
#chr1  3001631      3001632      100.000000   6      0
#chr1  3002177      3002178      100.000000   1      0
#chr1  3002338      3002339      100.000000   1      0
#chr1  3002386      3002387      75.000000    3      1
#chr1  3002599      3002600      100.000000   4      0
```

COMPUTATIONAL SCRIPTS 8-10: processing of Illumina RNA-seq reads

```
# This pipeline is for processing Illumina RNA-Sequencing reads (50bp SE
reads), sequenced on HiSeq2000. The goal is to obtain read counts for the
transcriptome.
## The pipeline uses a subset of samples from sequencing run 1, for
demonstration, and includes samples 3,8,17,22,26 and 29 corresponding to wild-
type samples Spg1, PL1, L1, Z1, P1 and D1, respectively.

### Script1: Concatenate all fastq.gz files and run FASTQ analysis

#!/bin/bash

# This script is used to generate concatenated fastq.gz files from original RNA
seq files
# working directory = /mnt/sequence/gaysinskaya/FACS_RNAseq1

FASTQ_DIR=/mnt/sequence/FastQ/130506_SN375_0223_BD25HHACXX/FastQ/Project_AB-
valeriyagaysinskaya/
SAMPLE="3 8 17 22 26 29"

mkdir -p rawdata
cd rawdata

###For all samples
for i in $SAMPLE
do
  echo "Concatenating Sample_$i "
  cat $FASTQ_DIR/Sample_$i/*.fastq.gz > Sample_$i\.fastq.gz
  fastqc Sample_$i\.fastq.gz &
done
```

****

```
### Script2: Align RNA-sequencing samples to mouse mm9 genome and iGenomes
transcriptome using Tophat program

#!/bin/bash

# This script is used to perform tophat on fastq.gz files
# working directory = /mnt/sequence/gaysinskaya/FACS_RNAseq1
# TopHat v2.0.7 is used

RAW_DIR=/mnt/sequence/gaysinskaya/FACS_RNAseq1/rawdata
TRANSCRIPTOME=/mnt/sequence/gaysinskaya/mm9-
genome/iGenome/Mus_musculus/UCSC/mm9/Annotation/Genes/genes.gtf
GENOME=/mnt/sequence/genomes/mouse/mm9

#Note, the annotation was downloaded from Illumina website, iGenomes collection
and dates to July,2015

#Sequencing run 1
SAMPLE="3 8 17 22 26 29"

mkdir -p tophat_iGenomes
cd tophat_iGenomes

###For all samples
for i in $SAMPLE
do
```

```
       nohup tophat -p 5 -o tophat$i -G $TRANSCRIPTOME $GENOME
$RAW_DIR/Sample_$i.fastq > itophat$i\_mm9.out
done
```

****

Script3: perform Htseq-count on sam files

```
#!/bin/bash

# This script "doHtseq-count.sh"is used to perform htseq-count on sam files
# working directory = /mnt/sequence/gaysinskaya/FACS_RNAseq1

# Make sure the sam file is in the directory from which htseq-count is executed
#Execute command like so: nohup ./doHtseq-count.sh > nohup_doHtseq&

DIR=/mnt/sequence/gaysinskaya/FACS_RNAseq1/tophat_iGenomes
GTF=/mnt/sequence/gaysinskaya/mm9-
genome/iGenome/Mus_musculus/UCSC/mm9/Annotation/Genes/genes.gtf

#Sequencing run 1
SAMPLE="3 8 17 22 26 29 31 32"

for i in $SAMPLE
do
       cd $DIR/tophat$i
       htseq-count -s no -a 10 $i\_accepted_hits.sam $GTF >
sample$i\_htseq.count
done

#the option -s <no> signifies that the data are not from a stranded protocol
and the -a option specifies a minimum score for the alignment quality.
```

Literature Cited

1. Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet. 2012;13(10):705-19. doi: 10.1038/nrg3273. PubMed PMID: 22986265.
2. Smallwood SA, Kelsey G. De novo DNA methylation: a germ cell perspective. Trends Genet. 2012;28(1):33-42. doi: 10.1016/j.tig.2011.09.004. PubMed PMID: 22019337.
3. Zamudio N, Bourc'his D. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity (Edinb). 2010;105(1):92-104. doi: 10.1038/hdy.2010.53. PubMed PMID: 20442734.
4. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 1997;13(8):335-40. PubMed PMID: 9260521.
5. Saitou M, Yamaji M. Primordial germ cells in mice. Cold Spring Harb Perspect Biol. 2012;4(11). doi: 10.1101/cshperspect.a008375. PubMed PMID: 23125014.
6. Gomperts M, Garcia-Castro M, Wylie C, Heasman J. Interactions between primordial germ cells play a role in their migration in mouse embryos. Development. 1994;120(1):135-41. PubMed PMID: 8119122.
7. Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res. 2013;23(4):616-27. doi: 10.1101/gr.148023.112. PubMed PMID: 23410886.
8. Tam PP, Snow MH. Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. J Embryol Exp Morphol. 1981;64:133-47. PubMed PMID: 7310300.
9. Manku G, Culty M. Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. Reproduction. 2015;149(3):R139-57. doi: 10.1530/REP-14-0431. PubMed PMID: 25670871.
10. Oakberg EF. Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. Am J Anat. 1956;99(3):507-16. doi: 10.1002/aja.1000990307. PubMed PMID: 13402729.
11. Russell LD. Histological and histopathological evaluation of the testis. 1st ed. Clearwater, Fl: Cache River Press; 1990. xiv, 286 p. p.
12. Hess RA, Renato de Franca L. Spermatogenesis and cycle of the seminiferous epithelium. Adv Exp Med Biol. 2008;636:1-15. doi: 10.1007/978-0-387-09597-4_1. PubMed PMID: 19856159.
13. Hogarth CA, Griswold MD. The key role of vitamin A in spermatogenesis. J Clin Invest. 2010;120(4):956-62. doi: 10.1172/JCI41303. PubMed PMID: 20364093.
14. Moens PB. Meiosis. Orlando: Academic Press; 1987. xii, 391 p. p.
15. Handel MA. Monitoring meiosis in gametogenesis. Theriogenology. 1998;49(2):423-30. PubMed PMID: 10732023.

16. Roth R. Chromosome replication during meiosis: identification of gene functions required for premeiotic DNA synthesis. Proc Natl Acad Sci U S A. 1973;70(11):3087-91. PubMed PMID: 4594035.
17. Callan HG. DNA replication in the chromosomes of eukaryotes. Cold Spring Harb Symp Quant Biol. 1974;38:195-203. PubMed PMID: 4598637.
18. Holm PB. The premeiotic DNA replication of euchromatin and heterochromatin in Lilium longiflorum (Thunb.). Carlsberg Research Communications. 1977;42(4):249-81. doi: 10.1007/bf02910453.
19. Cha RS, Weiner BM, Keeney S, Dekker J, Kleckner N. Progression of meiotic DNA replication is modulated by interchromosomal interaction proteins, negatively by Spo11p and positively by Rec8p. Genes Dev. 2000;14(4):493-503. PubMed PMID: 10691741.
20. Callan HG. Replication of DNA in eukaryotic chromosomes. Br Med Bull. 1973;29(3):192-5. PubMed PMID: 4807321.
21. Jaramillo-Lambert A, Ellefson M, Villeneuve AM, Engebrecht J. Differential timing of S phases, X chromosome replication, and meiotic prophase in the C. elegans germ line. Dev Biol. 2007;308(1):206-21. doi: 10.1016/j.ydbio.2007.05.019. PubMed PMID: 17599823.
22. Goren A, Cedar H. Replicating by the clock. Nat Rev Mol Cell Biol. 2003;4(1):25-32. doi: 10.1038/nrm1008. PubMed PMID: 12511866.
23. Rhind N, Gilbert DM. DNA replication timing. Cold Spring Harb Perspect Biol. 2013;5(8):a010132. doi: 10.1101/cshperspect.a010132. PubMed PMID: 23838440.
24. Lee B, Amon A. Meiosis: how to create a specialized cell cycle. Curr Opin Cell Biol. 2001;13(6):770-7. PubMed PMID: 11698195.
25. Page SL, Hawley RS. The genetics and molecular biology of the synaptonemal complex. Annu Rev Cell Dev Biol. 2004;20:525-58. doi: 10.1146/annurev.cellbio.19.111301.155141. PubMed PMID: 15473851.
26. Keeney S. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. Genome Dyn Stab. 2008;2:81-123. doi: 10.1007/7050_2007_026. PubMed PMID: 21927624.
27. Boateng KA, Bellani MA, Gregoretti IV, Pratto F, Camerini-Otero RD. Homologous pairing preceding SPO11-mediated double-strand breaks in mice. Dev Cell. 2013;24(2):196-205. Epub 2013/01/16. doi: 10.1016/j.devcel.2012.12.002. PubMed PMID: 23318132.
28. Gaysinskaya V, Soh IY, van der Heijden GW, Bortvin A. Optimized flow cytometry isolation of murine spermatocytes. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2014. Epub 2014/03/26. doi: 10.1002/cyto.a.22463. PubMed PMID: 24664803.
29. Grell RF, Oakberg EF, Generoso EE. Synaptonemal complexes at premeiotic interphase in the mouse spermatocyte. Proc Natl Acad Sci U S A. 1980;77(11):6720-3. PubMed PMID: 6935680.
30. Turner JM. Meiotic sex chromosome inactivation. Development. 2007;134(10):1823-31. doi: 10.1242/dev.000018. PubMed PMID: 17329371.
31. Brunner AM, Nanni P, Mansuy IM. Epigenetic marking of sperm by post-translational modification of histones and protamines. Epigenetics

Chromatin. 2014;7(1):2. doi: 10.1186/1756-8935-7-2. PubMed PMID: 24443974.
32. van der Heijden GW, Derijck AA, Posfai E, Giele M, Pelczar P, Ramos L, et al. Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. Nat Genet. 2007;39(2):251-8. doi: 10.1038/ng1949. PubMed PMID: 17237782.
33. Kierszenbaum AL, Tres LL. Nucleolar and perichromosomal RNA synthesis during meiotic prophase in the mouse testis. J Cell Biol. 1974;60(1):39-53. PubMed PMID: 4203361.
34. Szenker E, Ray-Gallet D, Almouzni G. The double face of the histone variant H3.3. Cell Res. 2011;21(3):421-34. doi: 10.1038/cr.2011.14. PubMed PMID: 21263457.
35. Tachibana M, Nozaki M, Takeda N, Shinkai Y. Functional dynamics of H3K9 methylation during meiotic prophase progression. EMBO J. 2007;26(14):3346-59. doi: 10.1038/sj.emboj.7601767. PubMed PMID: 17599069.
36. Takada Y, Naruse C, Costa Y, Shirakawa T, Tachibana M, Sharif J, et al. HP1gamma links histone methylation marks to meiotic synapsis in mice. Development. 2011;138(19):4207-17. doi: 10.1242/dev.064444. PubMed PMID: 21896631.
37. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell. 2001;107(3):323-37. PubMed PMID: 11701123.
38. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science. 2010;327(5967):836-40. doi: 10.1126/science.1183439. PubMed PMID: 20044539.
39. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6-21. doi: 10.1101/gad.947102. PubMed PMID: 11782440.
40. Bourc'his D, Proudhon C. Sexual dimorphism in parental imprint ontogeny and contribution to embryonic development. Mol Cell Endocrinol. 2008;282(1-2):87-94. doi: 10.1016/j.mce.2007.11.025. PubMed PMID: 18178305.
41. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993;366(6453):362-5. doi: 10.1038/366362a0. PubMed PMID: 8247133.
42. Antequera F. Structure, function and evolution of CpG island promoters. Cell Mol Life Sci. 2003;60(8):1647-58. doi: 10.1007/s00018-003-3088-6. PubMed PMID: 14504655.
43. Kass SU, Landsberger N, Wolffe AP. DNA methylation directs a time-dependent repression of transcription initiation. Curr Biol. 1997;7(3):157-65. PubMed PMID: 9395433.
44. Wade PA. Methyl CpG-binding proteins and transcriptional repression. Bioessays. 2001;23(12):1131-7. doi: 10.1002/bies.10008. PubMed PMID: 11746232.

45. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007;8(4):272-85. doi: 10.1038/nrg2072. PubMed PMID: 17363976.
46. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature. 2010;463(7284):1101-5. doi: 10.1038/nature08829. PubMed PMID: 20098412.
47. Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 2012;48(6):849-62. doi: 10.1016/j.molcel.2012.11.001. PubMed PMID: 23219530.
48. Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, et al. Programming and inheritance of parental DNA methylomes in mammals. Cell. 2014;157(4):979-91. doi: 10.1016/j.cell.2014.04.017. PubMed PMID: 24813617.
49. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. Nature. 2000;403(6769):501-2. doi: 10.1038/35000654. PubMed PMID: 10676950.
50. Inoue A, Matoba S, Zhang Y. Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation. Cell Res. 2012;22(12):1640-9. doi: 10.1038/cr.2012.160. PubMed PMID: 23184059.
51. Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res. 2011;21(12):1670-6. doi: 10.1038/cr.2011.189. PubMed PMID: 22124233.
52. Hajkova P. Epigenetic reprogramming in the germline: towards the ground state of the epigenome. Philos Trans R Soc Lond B Biol Sci. 2011;366(1575):2266-73. doi: 10.1098/rstb.2011.0042. PubMed PMID: 21727132.
53. Kubo N, Toh H, Shirane K, Shirakawa T, Kobayashi H, Sato T, et al. DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. BMC Genomics. 2015;16:624. doi: 10.1186/s12864-015-1833-5. PubMed PMID: 26290333.
54. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem. 2004;279(46):48350-9. doi: 10.1074/jbc.M403427200. PubMed PMID: 15339928.
55. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 2005;74:481-514. doi: 10.1146/annurev.biochem.74.010904.153721. PubMed PMID: 15952895.
56. La Salle S, Mertineit C, Taketo T, Moens PB, Bestor TH, Trasler JM. Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. Dev Biol. 2004;268(2):403-15. doi: 10.1016/j.ydbio.2003.12.031. PubMed PMID: 15063176.

57. Ravindran S. Barbara McClintock and the discovery of jumping genes. Proc Natl Acad Sci U S A. 2012;109(50):20198-9. doi: 10.1073/pnas.1219372109. PubMed PMID: 23236127.
58. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921. doi: 10.1038/35057062. PubMed PMID: 11237011.
59. Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev. 1999;9(6):657-63. PubMed PMID: 10607616.
60. Sookdeo A, Hepp CM, McClure MA, Boissinot S. Revisiting the evolution of mouse LINE-1 in the genomic era. Mob DNA. 2013;4(1):3. doi: 10.1186/1759-8753-4-3. PubMed PMID: 23286374.
61. Craig NL, Chandler M, Gellert M, Lambowitz A, Rice PA, Sandmeyer S. Mobile DNA III. Washington, DC: ASM Press; 2015. xxiv, 1321 pages p.
62. Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL, Moran JV. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr. 2015;3(2):MDNA3-0061-2014. doi: 10.1128/microbiolspec.MDNA3-0061-2014. PubMed PMID: 26104698.
63. Feschotte C. Transposable elements and the evolution of regulatory networks. Nat Rev Genet. 2008;9(5):397-405. doi: 10.1038/nrg2337. PubMed PMID: 18368054.
64. Makalowski W. Genomic scrap yard: how genomes utilize all that junk. Gene. 2000;259(1-2):61-7. PubMed PMID: 11163962.
65. Nekrutenko A, Li WH. Transposable elements are found in a large number of human protein-coding genes. Trends Genet. 2001;17(11):619-21. PubMed PMID: 11672845.
66. Jordan IK, Rogozin IB, Glazko GV, Koonin EV. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet. 2003;19(2):68-72. PubMed PMID: 12547512.
67. Pereira V, Enard D, Eyre-Walker A. The effect of transposable element insertions on gene expression evolution in rodents. PloS one. 2009;4(2):e4321. doi: 10.1371/journal.pone.0004321. PubMed PMID: 19183808.
68. Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420(6915):520-62. doi: 10.1038/nature01262. PubMed PMID: 12466850.
69. Goodier JL, Ostertag EM, Du K, Kazazian HH, Jr. A novel active L1 retrotransposon subfamily in the mouse. Genome Res. 2001;11(10):1677-85. doi: 10.1101/gr.198301. PubMed PMID: 11591644.
70. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 2004;428(6982):493-521. doi: 10.1038/nature02426. PubMed PMID: 15057822.
71. Kazazian HH, Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences

represents a novel mechanism for mutation in man. Nature. 1988;332(6160):164-6. doi: 10.1038/332164a0. PubMed PMID: 2831458.
72. Belancio VP, Deininger PL, Roy-Engel AM. LINE dancing in the human genome: transposable elements and disease. Genome Med. 2009;1(10):97. doi: 10.1186/gm97. PubMed PMID: 19863772.
73. Miousse IR, Koturbash I. The Fine LINE: Methylation Drawing the Cancer Landscape. Biomed Res Int. 2015;2015:131547. doi: 10.1155/2015/131547. PubMed PMID: 26448926.
74. Molaro A, Falciatori I, Hodges E, Aravin AA, Marran K, Rafii S, et al. Two waves of de novo methylation during mouse germ cell development. Genes Dev. 2014;28(14):1544-9. doi: 10.1101/gad.244350.114. PubMed PMID: 25030694.
75. Pezic D, Manakov SA, Sachidanandam R, Aravin AA. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes Dev. 2014;28(13):1410-28. doi: 10.1101/gad.240895.114. PubMed PMID: 24939875.
76. Aravin AA, Bourc'his D. Small RNA guides for de novo DNA methylation in mammalian germ cells. Genes Dev. 2008;22(8):970-5. doi: 10.1101/gad.1669408. PubMed PMID: 18413711.
77. Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N, et al. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev. 2015;29(12):1256-70. doi: 10.1101/gad.257840.114. PubMed PMID: 26109049.
78. Barroca V, Lassalle B, Coureuil M, Louis JP, Le Page F, Testart J, et al. Mouse differentiating spermatogonia can generate germinal stem cells in vivo. Nat Cell Biol. 2009;11(2):190-6. Epub 2008/12/23. doi: 10.1038/ncb1826. PubMed PMID: 19098901.
79. Bastos H, Lassalle B, Chicheportiche A, Riou L, Testart J, Allemand I, et al. Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2005;65(1):40-9. Epub 2005/03/22. doi: 10.1002/cyto.a.20129. PubMed PMID: 15779065.
80. Lassalle B, Bastos H, Louis JP, Riou L, Testart J, Dutrillaux B, et al. 'Side Population' cells in adult mouse testis express Bcrp1 gene and are enriched in spermatogonia and germinal stem cells. Development. 2004;131(2):479-87. Epub 2003/12/19. doi: 10.1242/dev.00918. PubMed PMID: 14681185.
81. Fallahi M, Getun, IV, Wu ZK, Bois, PR. A Global Expression Switch Marks Pachytene Initiation during Mouse Male Meiosis. Genes. 2010;1(3):469-83. Epub 2010. doi: 10.3390/genes1030469.
82. Getun IV, Torres B, Bois PR. Flow cytometry purification of mouse meiotic cells. J Vis Exp. 2011;(50). Epub 2011/04/29. doi: 10.3791/2602. PubMed PMID: 21525843.
83. Getun IV, Wu ZK, Khalil AM, Bois PR. Nucleosome occupancy landscape and dynamics at mouse recombination hotspots. EMBO reports. 2010;11(7):555-60. Epub 2010/05/29. doi: 10.1038/embor.2010.79. PubMed PMID: 20508641.

84. Zickler D, Kleckner N. The leptotene-zygotene transition of meiosis. Annu Rev Genet. 1998;32:619-97. Epub 1999/02/03. doi: 10.1146/annurev.genet.32.1.619. PubMed PMID: 9928494.
85. Scherthan H, Weich S, Schwegler H, Heyting C, Harle M, Cremer T. Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. J Cell Biol. 1996;134(5):1109-25. Epub 1996/09/01. PubMed PMID: 8794855.
86. Bass HW. Telomere dynamics unique to meiotic prophase: formation and significance of the bouquet. Cell Mol Life Sci. 2003;60(11):2319-24. Epub 2003/11/20. doi: 10.1007/s00018-003-3312-4. PubMed PMID: 14625678.
87. Scherthan H. A bouquet makes ends meet. Nat Rev Mol Cell Biol. 2001;2(8):621-7. Epub 2001/08/03. doi: 10.1038/35085086. PubMed PMID: 11483995.
88. Nabeshima K. Collaborative homologous pairing during C. elegans meiosis. Worm. 2012;1(2):116-20. Epub 2012/04/01. doi: 10.4161/worm.19528. PubMed PMID: 24058834.
89. Rog O, Dernburg AF. Chromosome pairing and synapsis during Caenorhabditis elegans meiosis. Curr Opin Cell Biol. 2013;25(3):349-56. Epub 2013/04/13. doi: 10.1016/j.ceb.2013.03.003. PubMed PMID: 23578368.
90. Burgess SM. Homologous chromosome associations and nuclear order in meiotic and mitotically dividing cells of budding yeast. Adv Genet. 2002;46:49-90. Epub 2002/04/05. PubMed PMID: 11931237.
91. Ding DQ, Haraguchi T, Hiraoka Y. From meiosis to postmeiotic events: alignment and recognition of homologous chromosomes in meiosis. FEBS J. 2010;277(3):565-70. Epub 2009/12/18. doi: 10.1111/j.1742-4658.2009.07501.x. PubMed PMID: 20015081.
92. Ding DQ, Okamasa K, Yamane M, Tsutsumi C, Haraguchi T, Yamamoto M, et al. Meiosis-specific noncoding RNA mediates robust pairing of homologous chromosomes in meiosis. Science. 2012;336(6082):732-6. Epub 2012/05/15. doi: 10.1126/science.1219518. PubMed PMID: 22582262.
93. Chowdhury R, Bois PR, Feingold E, Sherman SL, Cheung VG. Genetic analysis of variation in human meiotic recombination. PLoS Genet. 2009;5(9):e1000648. Epub 2009/09/19. doi: 10.1371/journal.pgen.1000648. PubMed PMID: 19763160.
94. Di Giacomo M, Comazzetto S, Saini H, De Fazio S, Carrieri C, Morgan M, et al. Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. Mol Cell. 2013;50(4):601-8. Epub 2013/05/28. doi: 10.1016/j.molcel.2013.04.026. PubMed PMID: 23706823.
95. Falciatori I, Borsellino G, Haliassos N, Boitani C, Corallini S, Battistini L, et al. Identification and enrichment of spermatogonial stem cells displaying side-population phenotype in immature mouse testis. FASEB J. 2004;18(2):376-8. Epub 2003/12/23. doi: 10.1096/fj.03-0744fje. PubMed PMID: 14688197.
96. Shimizu Y, Motohashi N, Iseki H, Kunita S, Sugiyama F, Yagami K. A novel subpopulation lacking Oct4 expression in the testicular side population. Int J Mol Med. 2006;17(1):21-8. Epub 2005/12/06. PubMed PMID: 16328007.

97. Yoshida S, Sukeno M, Nakagawa T, Ohbo K, Nagamatsu G, Suda T, et al. The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development. 2006;133(8):1495-505. Epub 2006/03/17. doi: 10.1242/dev.02316. PubMed PMID: 16540512.
98. Kluin PM, Kramer MF, de Rooij DG. Spermatogenesis in the immature mouse proceeds faster than in the adult. International journal of andrology. 1982;5(3):282-94. Epub 1982/06/01. PubMed PMID: 7118267.
99. Mori C, Nakamura N, Dix DJ, Fujioka M, Nakagawa S, Shiota K, et al. Morphological analysis of germ cell apoptosis during postnatal testis development in normal and Hsp 70-2 knockout mice. Dev Dyn. 1997;208(1):125-36. Epub 1997/01/01. doi: 10.1002/(SICI)1097-0177(199701)208:1<125::AID-AJA12>3.0.CO;2-5. PubMed PMID: 8989527.
100. Peters AH, Plug AW, van Vugt MJ, de Boer P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology. 1997;5(1):66-8. Epub 1997/02/01. PubMed PMID: 9088645.
101. Golebiewska A, Brons NH, Bjerkvig R, Niclou SP. Critical appraisal of the side population assay in stem cell and cancer stem cell research. Cell Stem Cell. 2011;8(2):136-47. Epub 2011/02/08. doi: 10.1016/j.stem.2011.01.007. PubMed PMID: 21295271.
102. Rodriguez-Casuriaga R, Geisinger A, Santinaque FF, Lopez-Carro B, Folle GA. High-purity flow sorting of early meiocytes based on DNA analysis of guinea pig spermatogenic cells. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2011;79(8):625-34. Epub 2011/04/27. doi: 10.1002/cyto.a.21067. PubMed PMID: 21520399.
103. Givan AL. Flow cytometry: an introduction. Methods in molecular biology. 2011;699:1-29. Epub 2010/12/01. doi: 10.1007/978-1-61737-950-5_1. PubMed PMID: 21116976.
104. Ding X, Xu R, Yu J, Xu T, Zhuang Y, Han M. SUN1 is required for telomere attachment to nuclear envelope and gametogenesis in mice. Dev Cell. 2007;12(6):863-72. Epub 2007/06/05. doi: 10.1016/j.devcel.2007.03.018. PubMed PMID: 17543860.
105. Ishiguro K, Kim J, Fujiyama-Nakamura S, Kato S, Watanabe Y. A new meiosis-specific cohesin complex implicated in the cohesin code for homologous pairing. EMBO reports. 2011;12(3):267-75. Epub 2011/01/29. doi: 10.1038/embor.2011.2. PubMed PMID: 21274006.
106. Liebe B, Petukhova G, Barchi M, Bellani M, Braselmann H, Nakano T, et al. Mutations that affect meiosis in male mice influence the dynamics of the mid-preleptotene and bouquet stages. Exp Cell Res. 2006;312(19):3768-81. Epub 2006/10/03. doi: 10.1016/j.yexcr.2006.07.019. PubMed PMID: 17010969.
107. Pellegrini M, Di Siena S, Claps G, Di Cesare S, Dolci S, Rossi P, et al. Microgravity promotes differentiation and meiotic entry of postnatal mouse male germ cells. PloS one. 2010;5(2):e9064. Epub 2010/02/09. doi: 10.1371/journal.pone.0009064. PubMed PMID: 20140225.

108. Inagaki A, Schoenmakers S, Baarends WM. DNA double strand break repair, chromosome synapsis and transcriptional silencing in meiosis. Epigenetics : official journal of the DNA Methylation Society. 2010;5(4):255-66. Epub 2010/04/07. PubMed PMID: 20364103.
109. Ellwart JW, Dormer P. Vitality measurement using spectrum shift in Hoechst 33342 stained cells. Cytometry. 1990;11(2):239-43. doi: 10.1002/cyto.990110204. PubMed PMID: 1690626.
110. Pedersen RA, Schatten GP, Handel MA. Meiosis and Gametogenesis: Academic Press; 1997.
111. Evenson D, Darzynkiewicz Z, Jost L, Janca F, Ballachey B. Changes in accessibility of DNA to various fluorochromes during spermatogenesis. Cytometry. 1986;7(1):45-53. doi: 10.1002/cyto.990070107. PubMed PMID: 2419055.
112. Ohta H, Sakaide Y, Wakayama T. Generation of progeny via ICSI following enrichment of elongated spermatids from mouse testis by flow-cytometric cell sorting. Human reproduction. 2007;22(6):1612-6. doi: 10.1093/humrep/dem064. PubMed PMID: 17428882.
113. Davies D. Cell separations by flow cytometry. Methods in molecular biology. 2012;878:185-99. doi: 10.1007/978-1-61779-854-2_12. PubMed PMID: 22674134.
114. Givan AL. The Sorting of Cells. Flow Cytometry: First Principles. 2nd ed. New York: Wiley-Liss; 2001. p. 159.
115. Herzenberg LA, Tung J, Moore WA, Herzenberg LA, Parks DR. Interpreting flow cytometry data: a guide for the perplexed. Nat Immunol. 2006;7(7):681-5. doi: 10.1038/ni0706-681. PubMed PMID: 16785881.
116. Soper SF, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de Boer P, et al. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. Dev Cell. 2008;15(2):285-97. doi: 10.1016/j.devcel.2008.05.015. PubMed PMID: 18694567.
117. Petersen TW, Ibrahim SF, Diercks AH, van den Engh G. Chromatic shifts in the fluorescence emitted by murine thymocytes stained with Hoechst 33342. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2004;60(2):173-81. doi: 10.1002/cyto.a.20058. PubMed PMID: 15290718.
118. Darzynkiewicz Z. Critical Aspects in Analysis of Cellular DNA Content. Current Protocols in Cytometry. 56: John Wiley & Sons, Inc.; 2011. p. 7.2.1-7.2.8.
119. Matson CK, Murphy MW, Griswold MD, Yoshida S, Bardwell VJ, Zarkower D. The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. Dev Cell. 2010;19(4):612-24. doi: 10.1016/j.devcel.2010.09.010. PubMed PMID: 20951351.
120. Zhang T, Murphy MW, Gearhart MD, Bardwell VJ, Zarkower D. The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis.

Development. 2014;141(19):3662-71. doi: 10.1242/dev.113936. PubMed PMID: 25249458.
121. Yuen BT, Bush KM, Barrilleaux BL, Cotterman R, Knoepfler PS. Histone H3.3 regulates dynamic chromatin states during spermatogenesis. Development. 2014;141(18):3483-94. doi: 10.1242/dev.106450. PubMed PMID: 25142466.
122. Petriz J. Flow cytometry of the side population (SP). Current protocols in cytometry / editorial board, J Paul Robinson, managing editor [et al]. 2013;Chapter 9:Unit9 23. doi: 10.1002/0471142956.cy0923s64. PubMed PMID: 23546779.
123. Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. Nat Rev Genet. 2009;10(11):805-11. doi: 10.1038/nrg2651. PubMed PMID: 19789556.
124. Bestor TH. The DNA methyltransferases of mammals. Hum Mol Genet. 2000;9(16):2395-402. PubMed PMID: 11005794.
125. Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol. 2001;309(5):1189-99. doi: 10.1006/jmbi.2001.4709. PubMed PMID: 11399088.
126. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science. 2001;293(5532):1089-93. doi: 10.1126/science.1063443. PubMed PMID: 11498579.
127. Aravin AA, van der Heijden GW, Castaneda J, Vagin VV, Hannon GJ, Bortvin A. Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. PLoS Genet. 2009;5(12):e1000764. doi: 10.1371/journal.pgen.1000764. PubMed PMID: 20011505.
128. Seisenberger S, Peat JR, Hore TA, Santos F, Dean W, Reik W. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. Philos Trans R Soc Lond B Biol Sci. 2013;368(1609):20110330. doi: 10.1098/rstb.2011.0330. PubMed PMID: 23166394.
129. Bortvin A. PIWI-interacting RNAs (piRNAs) - a mouse testis perspective. Biochemistry (Mosc). 2013;78(6):592-602. doi: 10.1134/S0006297913060059. PubMed PMID: 23980886.
130. Branciforte D, Martin SL. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol Cell Biol. 1994;14(4):2584-92. PubMed PMID: 8139560.
131. Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. Nature. 2011;480(7376):264-7. doi: 10.1038/nature10672. PubMed PMID: 22121019.
132. Pandey RR, Tokuzawa Y, Yang Z, Hayashi E, Ichisaka T, Kajita S, et al. Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. Proc Natl Acad Sci U S A. 2013;110(41):16492-7. doi: 10.1073/pnas.1316316110. PubMed PMID: 24067652.

133. Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y, et al. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev Cell. 2009;17(6):775-87. doi: 10.1016/j.devcel.2009.10.012. PubMed PMID: 20059948.
134. van der Heijden GW, Bortvin A. Transient relaxation of transposon silencing at the onset of mammalian meiosis. Epigenetics : official journal of the DNA Methylation Society. 2009;4(2):76-9. PubMed PMID: 19252417.
135. Castaneda J, Genzor P, van der Heijden GW, Sarkeshik A, Yates JR, 3rd, Ingolia NT, et al. Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. EMBO J. 2014;33(18):1999-2019. doi: 10.15252/embj.201386855. PubMed PMID: 25063675.
136. Adler ID. Comparison of the duration of spermatogenesis between male rodents and humans. Mutat Res. 1996;352(1-2):169-72. doi: 10.1016/0027-5107(95)00223-5. PubMed PMID: 8676906.
137. Lam I, Keeney S. Mechanism and regulation of meiotic recombination initiation. Cold Spring Harb Perspect Biol. 2015;7(1):a016634. doi: 10.1101/cshperspect.a016634. PubMed PMID: 25324213.
138. Baudat F, Imai Y, de Massy B. Meiotic recombination in mammals: localization and regulation. Nat Rev Genet. 2013;14(11):794-806. doi: 10.1038/nrg3573. PubMed PMID: 24136506.
139. Zickler D, Kleckner N. Meiotic chromosomes: integrating structure and function. Annu Rev Genet. 1999;33:603-754. doi: 10.1146/annurev.genet.33.1.603. PubMed PMID: 10690419.
140. Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96-9. doi: 10.1038/nature02886. PubMed PMID: 15318244.
141. Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, et al. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell. 2007;12(4):503-14. doi: 10.1016/j.devcel.2007.03.001. PubMed PMID: 17395546.
142. Martin SL, Branciforte D. Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. Mol Cell Biol. 1993;13(9):5383-92. PubMed PMID: 8395003.
143. Lambrot R, Jones S, Saint-Phar S, Kimmins S. Specialized distribution of the histone methyltransferase Ezh2 in the nuclear apical region of round spermatids and its interaction with the histone variant H1t2. J Androl. 2012;33(5):1058-66. doi: 10.2164/jandrol.111.013870. PubMed PMID: 22323620.
144. Xiao N, Kam C, Shen C, Jin W, Wang J, Lee KM, et al. PICK1 deficiency causes male infertility in mice by disrupting acrosome formation. J Clin Invest. 2009;119(4):802-12. doi: 10.1172/JCI36230. PubMed PMID: 19258705.
145. Mahadevaiah SK, Turner JM, Baudat F, Rogakou EP, de Boer P, Blanco-Rodriguez J, et al. Recombinational DNA double-strand breaks in mice precede synapsis. Nat Genet. 2001;27(3):271-6. doi: 10.1038/85830. PubMed PMID: 11242108.

146. Gaysinskaya V, Bortvin A. Flow cytometry of murine spermatocytes. Current protocols in cytometry / editorial board, J Paul Robinson, managing editor [et al]. 2015;72:7 44 1-7 24. doi: 10.1002/0471142956.cy0744s72. PubMed PMID: 25827485.
147. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571-2. doi: 10.1093/bioinformatics/btr167. PubMed PMID: 21493656.
148. Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing data. Nat Methods. 2012;9(2):145-51. doi: 10.1038/nmeth.1828. PubMed PMID: 22290186.
149. Pope BD, Chandra T, Buckley Q, Hoare M, Ryba T, Wiseman FK, et al. Replication-timing boundaries facilitate cell-type and species-specific regulation of a rearranged human chromosome in mouse. Hum Mol Genet. 2012;21(19):4162-70. doi: 10.1093/hmg/dds232. PubMed PMID: 22736031.
150. Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. 2010;20(6):761-70. doi: 10.1101/gr.099655.109. PubMed PMID: 20430782.
151. Yaffe E, Farkash-Amar S, Polten A, Yakhini Z, Tanay A, Simon I. Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. PLoS Genet. 2010;6(7):e1001011. doi: 10.1371/journal.pgen.1001011. PubMed PMID: 20617169.
152. Weddington N, Stuy A, Hiratani I, Ryba T, Yokochi T, Gilbert DM. ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. BMC Bioinformatics. 2008;9:530. doi: 10.1186/1471-2105-9-530. PubMed PMID: 19077204.
153. Koren A, Handsaker RE, Kamitaki N, Karlic R, Ghosh S, Polak P, et al. Genetic variation in human DNA replication timing. Cell. 2014;159(5):1015-26. doi: 10.1016/j.cell.2014.10.025. PubMed PMID: 25416942.
154. Yarosh W, Spradling AC. Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells. Genes Dev. 2014;28(16):1840-55. doi: 10.1101/gad.245811.114. PubMed PMID: 25128500.
155. Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J Cell Biol. 2004;166(4):493-505. doi: 10.1083/jcb.200403109. PubMed PMID: 15302854.
156. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. 2012;13(10):R83. doi: 10.1186/gb-2012-13-10-r83. PubMed PMID: 23034175.
157. Razin A, Cedar H. DNA methylation and gene expression. Microbiol Rev. 1991;55(3):451-8. PubMed PMID: 1943996.
158. Payer B, Lee JT. X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet. 2008;42:733-72. doi: 10.1146/annurev.genet.42.110807.091711. PubMed PMID: 18729722.

159. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9. doi: 10.1093/bioinformatics/btu638. PubMed PMID: 25260700.
160. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105-11. doi: 10.1093/bioinformatics/btp120. PubMed PMID: 19289445.
161. Rabionet M, Bayerle A, Jennemann R, Heid H, Fuchser J, Marsching C, et al. Male meiotic cytokinesis requires ceramide synthase 3-dependent sphingolipids with unique membrane anchors. Hum Mol Genet. 2015;24(17):4792-808. doi: 10.1093/hmg/ddv204. PubMed PMID: 26045466.
162. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. 2013;8(9):1765-86. doi: 10.1038/nprot.2013.099. PubMed PMID: 23975260.
163. O'Shaughnessy PJ, Willerton L, Baker PJ. Changes in Leydig cell gene expression during development in the mouse. Biol Reprod. 2002;66(4):966-75. PubMed PMID: 11906915.
164. Laiho A, Kotaja N, Gyenesei A, Sironen A. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. PloS one. 2013;8(4):e61558. doi: 10.1371/journal.pone.0061558. PubMed PMID: 23613874.
165. Hammoud SS, Low DH, Yi C, Lee CL, Oatley JM, Payne CJ, et al. Transcription and imprinting dynamics in developing postnatal male germline stem cells. Genes Dev. 2015;29(21):2312-24. doi: 10.1101/gad.261925.115. PubMed PMID: 26545815.
166. Handel MA, Schimenti JC. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat Rev Genet. 2010;11(2):124-36. doi: 10.1038/nrg2723. PubMed PMID: 20051984.
167. Anderson EL, Baltus AE, Roepers-Gajadien HL, Hassold TJ, de Rooij DG, van Pelt AM, et al. Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. Proc Natl Acad Sci U S A. 2008;105(39):14976-80. doi: 10.1073/pnas.0807297105. PubMed PMID: 18799751.
168. Sun R, Qi H. Dynamic expression of combinatorial replication-dependent histone variant genes during mouse spermatogenesis. Gene Expr Patterns. 2014;14(1):30-41. doi: 10.1016/j.gep.2013.10.002. PubMed PMID: 24140495.
169. Tang MC, Jacobs SA, Mattiske DM, Soh YM, Graham AN, Tran A, et al. Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. PLoS Genet. 2015;11(2):e1004964. doi: 10.1371/journal.pgen.1004964. PubMed PMID: 25675407.
170. Modzelewski AJ, Holmes RJ, Hilz S, Grimson A, Cohen PE. AGO4 regulates entry into meiosis and influences silencing of sex chromosomes in the male mouse germline. Dev Cell. 2012;23(2):251-64. doi: 10.1016/j.devcel.2012.07.003. PubMed PMID: 22863743.

171. Shima JE, McLean DJ, McCarrey JR, Griswold MD. The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. Biol Reprod. 2004;71(1):319-30. doi: 10.1095/biolreprod.103.026880. PubMed PMID: 15028632.
172. Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. 2013;3(6):2179-90. doi: 10.1016/j.celrep.2013.05.031. PubMed PMID: 23791531.
173. Margolin G, Khil PP, Kim J, Bellani MA, Camerini-Otero RD. Integrated transcriptome analysis of mouse spermatogenesis. BMC Genomics. 2014;15:39. doi: 10.1186/1471-2164-15-39. PubMed PMID: 24438502.
174. Eddy EM, O'Brien DA. Gene expression during mammalian meiosis. Curr Top Dev Biol. 1998;37:141-200. PubMed PMID: 9352186.
175. Gan H, Wen L, Liao S, Lin X, Ma T, Liu J, et al. Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. Nat Commun. 2013;4:1995. doi: 10.1038/ncomms2995. PubMed PMID: 23759713.
176. Sharif J, Koseki H. Recruitment of Dnmt1 roles of the SRA protein Np95 (Uhrf1) and other factors. Prog Mol Biol Transl Sci. 2011;101:289-310. doi: 10.1016/B978-0-12-387685-0.00008-1. PubMed PMID: 21507355.
177. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007;450(7171):908-12. doi: 10.1038/nature06397. PubMed PMID: 17994007.
178. Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317(5845):1760-4. doi: 10.1126/science.1147939. PubMed PMID: 17673620.
179. Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem. 2004;279(26):27816-23. doi: 10.1074/jbc.M400181200. PubMed PMID: 15105426.
180. Bestor TH, Edwards JR, Boulard M. Notes on the role of dynamic DNA methylation in mammalian development. Proc Natl Acad Sci U S A. 2015;112(22):6796-9. doi: 10.1073/pnas.1415301111. PubMed PMID: 25368180.
181. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009;1(2):239-59. doi: 10.2217/epi.09.33. PubMed PMID: 20495664.
182. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33 Suppl:245-54. doi: 10.1038/ng1089. PubMed PMID: 12610534.
183. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454(7205):766-70. doi: 10.1038/nature07107. PubMed PMID: 18600261.
184. Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N. Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC

Genomics. 2014;15:583. doi: 10.1186/1471-2164-15-583. PubMed PMID: 25012247.
185. Kappler JW. The kinetics of DNA methylation in cultures of a mouse adrenal cell line. Journal of cellular physiology. 1970;75(1):21-31. doi: 10.1002/jcp.1040750104.
186. Gruenbaum Y, Szyf M, Cedar H, Razin A. Methylation of replicating and post-replicated mouse L-cell DNA. Proceedings of the National Academy of Sciences of the United States of America. 1983;80(16):4919-21.
187. Vandiver AR, Idrizi A, Rizzardi L, Feinberg AP, Hansen KD. DNA methylation is stable during replication and cell cycle arrest. Sci Rep. 2015;5:17911. doi: 10.1038/srep17911. PubMed PMID: 26648411.
188. Lin DW, Chung BP, Kaiser P. S-adenosylmethionine limitation induces p38 mitogen-activated protein kinase and triggers cell cycle arrest in G1. J Cell Sci. 2014;127(Pt 1):50-9. doi: 10.1242/jcs.127811. PubMed PMID: 24155332.
189. Li XZ, Roy CK, Dong X, Bolcun-Filas E, Wang J, Han BW, et al. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell. 2013;50(1):67-81. doi: 10.1016/j.molcel.2013.02.016. PubMed PMID: 23523368.
190. Braun RE. Post-transcriptional control of gene expression during spermatogenesis. Semin Cell Dev Biol. 1998;9(4):483-9. doi: 10.1006/scdb.1998.0226. PubMed PMID: 9813196.
191. Cha RS, Weiner BM, Keeney S, Dekker J, Kleckner N. Progression of meiotic DNA replication is modulated by interchromosomal interaction proteins, negatively by Spo11p and positively by Rec8p. Genes Dev. 2000;14(4):493-503.
192. Zickler D, Kleckner N. Meiotic chromosomes: integrating structure and function. Meiotic chromosomes: integrating structure and function. 1999. doi: 10.1146/annurev.genet.33.1.603.
193. Maloisel L, Rossignol JL. Suppression of crossing-over by DNA methylation in Ascobolus. Genes Dev. 1998;12(9):1381-9. PubMed PMID: 9573054.
194. Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, et al. Epigenetic Remodeling of Meiotic Crossover Frequency in Arabidopsis thaliana DNA Methyltransferase Mutants. PLoS Genet. 2012;8(8). doi: 10.1371/journal.pgen.1002844.
195. Yelina NE, Lambing C, Hardcastle TJ, Zhao X, Santos B, Henderson IR. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes Dev. 2015;29(20):2183-202. doi: 10.1101/gad.270876.115.
196. Robinson WP, Lalande M. Sex-specific meiotic recombination in the Prader--Willi/Angelman syndrome imprinted region. Hum Mol Genet. 1995;4(5):801-6. PubMed PMID: 7633438.
197. Malki S, van der Heijden GW, O'Donnell KA, Martin SL, Bortvin A. A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. Dev Cell. 2014;29(5):521-33. doi: 10.1016/j.devcel.2014.04.027. PubMed PMID: 24882376.

198. Hayatsu H. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis--a personal account. Proc Jpn Acad Ser B Phys Biol Sci. 2008;84(8):321-30. PubMed PMID: 18941305.
199. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. 2011;43(8):768-75. doi: 10.1038/ng.865. PubMed PMID: 21706001.
200. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-78. doi: 10.1038/nprot.2012.016. PubMed PMID: 22383036.
201. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40(10):4288-97. doi: 10.1093/nar/gks042. PubMed PMID: 22287627.

BIOGRAPHICAL SKETCH

NAME: Gaysinskaya, Valeriya

POSITION TITLE: Predoctoral Fellow

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	Completion Date	FIELD OF STUDY
Hunter College, New York, NY	B.A.	2005	Biology
Johns Hopkins University/Carnegie Institution, Baltimore, MD	Ph.D.	2016	Cell, Molecular and Developmental biology

A. Personal Statement

Prior to my graduate work, I did research in three major areas of biology: (1) developmental biology, studying aspects of polar body formation in a non-model organism, clam *Spisula solidissima*, (2) neurobiology, focusing on the effects of maternal high-fat diet on brain development of the offspring in rat and (3) virology, working on developing mouse model for hepatitis C. In my graduate studies I investigated the developmental biology and DNA epigenetics of mouse germ cells and explored the biology of the transposable elements - the highly abundant remnants of ancient infections, in mouse germline. My thesis research focused on characterizing and understanding a unique aspect of meiosis, namely, transient relaxation of DNA methylation at meiotic onset. My work highlights a period in spermatogenesis when genome-wide DNA methylation, including DNA methylation-dependent control of potentially dangerous transposable elements, is relaxed. My work has an important implication for understanding sensitivity of meiotic germ cells to retrotransposon expression, implicating the epigenetically relaxed state of meiotic onset in this sensitivity. My thesis research enhances our understanding of the epigenetic dynamics and plasticity of mouse male germ cells during meiotic prophase I. My expertise covers unique skills, like fluorescence activating cell sorting of spermatocytes, bioinformatics skills, including analysis of genome-wide DNA methylation and transcription post deep sequencing and common methods in cell and molecular biology, such as immunohistochemistry and microscopy.

B. Positions and Honors

Positions and Employment

2009- Predoc fellow. Johns Hopkins University/Carnegie Institution. Lab of Dr. Alex Bortvin.

2007-2009 Lab technician. The Rockefeller University. Lab of Dr. Charles Rice.

2004-2007 Volunteer/Research Assistant. The Rockefeller University. Lab of Dr. Sarah Leibowitz.

2003-2003 HHMI Undergraduate fellow. Marine Biological Institute. Lab of Dr. William D. Cohen.

Honors

2013 The J. Brien Key Graduate Travel Award

C. Contribution to Science

1. My earliest work contributed to our understanding of the formation and function of a polar body. A polar body is a byproduct of oocyte meiosis and it is a fundamental feature of sexual reproduction. This work describes the sequence, spatial organization and mechanism of major cytoskeletal events contributing to unequal cytokinesis and polar body formation in clam *Spisula solidissima.*

a. Pielak RM, **Gaysinskaya VA**, Cohen WD. Formation and function of the polar body contractile ring in Spisula. Developmental Biology. 2004;269(2):421-32
b. R.M.Pielak, **V.A. Gaysinskaya**, and W.D.Cohen. "Cytoskeletal events preceding polar body formation in activated *Spisula* eggs". Biological Bulletin. 2003;205(2):192-3.

2. In the field of neurobiology I worked on describing the effects of overeating and obesity on rat brain, particularly, the hypothalamic region. I was involved in efforts showing that maternal high-fat diet affects fetal brain development.

a. Chang GQ, **Gaysinskaya V**, Karatayev O, Leibowitz SF. Maternal high-fat diet and fetal programming: increased proliferation of hypothalamic peptide-producing neurons that increase risk for overeating and obesity. J Neurosci. 2008;28(46):12107-19.
b. **Gaysinskaya VA**, Karatayev O, Chang GQ, Leibowitz SF. Increased caloric intake after a high-fat preload: relation to circulating triglycerides and orexigenic peptides. Physiology and Behavior. 2007;91(1):142-53.
c. G.-Q. Chang, O. Karatayev, R. Ahsan, **V. Gaysinskaya**, Z. Marwil, and S.F. Leibowitz. "Dietary fat stimulates endogenous enkephalin and dynorphin in the paraventricular nucleus: role of circulating triglycerides. American Journal of Physiology- Endocrinology and Metabolism 2007;292(2):E561-70.

3. Interest in disease-causing pathogens led me to work in the field of infections dieseases. Among a team of viologists and molecular biologists, I worked on developing humanized mice model for hepatitis C. This work enabled modeling of human hepatitis C in mice and elucidation of the mechanism and the pathology of this worldwide viral infection.

a. Ploss A, Khetani SR, Jones CT, Syder AJ, Trehan K, **Gaysinskaya VA**, Mu K, Ritola K, Rice CM, Bhatia SN. Persistent hepatitis C virus infection in microscale primary human hepatocyte cultures. Proc Natl Acad Sci U S A. 2010;107(7):3141-5.
b. Ploss A, Evans MJ, **Gaysinskaya VA**, Panis M, You H, de Jong YP, Rice CM. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. Nature. 2009;457(7231):882-6.

4. In my graduate research, I focused on understanding epigenetic changes in meiosis. I describe previously uncharacterized transient relaxation in DNA methylation at the onset of meiosis in mouse male germ cells. These studies demonstrate that relaxation in DNA methylation is linked to pre-meiotic DNA replication and results in transient relaxation of transposable element silencing in meiosis. My work identifies an additional step in epigenetic programming of the male germline that may be important for gamete quality control.

a. **Valeriya Gaysinskaya,** Godfried van der Heijden, Brendan Miller, Kasper D. Hansen, Alex Bortvin. Transient relaxation of DNA methylation at the onset of meiosis in male mice. *Submitted*.
b. **Gaysinskaya V** and Bortvin A. Flow cytometry of murine spermatocytes. Curr Protoc Cytom. 2015;72(7):7.44.1-24.
c. **Gaysinskaya V**, Soh IY, van der Heijden GW, Bortvin A. Optimized flow cytometry isolation of murine spermatocytes. Cytometry A. 2014;85(6):556-65.

5. During my graduate training I became interested in issues related to bioethics and the politics of public health. I explored a particularly pertinent issue of mitorchondrial replacement therapy in the context of three-parent in vitro fertilization (IVF).

a. **Valeriya Gaysinskaya.** "The many names of mitochondrial transfer IVF." March 13, 2015. Johns Hopkins Berman Institute of Bioethics, bioethicsbulletin.org.
b. **Valeriya Gaysinskaya.** "Replacing faulty mitochondria". February 3, 2015. Johns Hopkins Berman Institute of Bioethics, bioethicsbulletin.org.